

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
24 October 2002 (24.10.2002)

PCT

(10) International Publication Number  
WO 02/083855 A2

- (51) International Patent Classification<sup>7</sup>: **C12N** **Lawrence** [US/US]; 9 Glenley Terrace, Brighton, MA 02135 (US).
- (21) International Application Number: PCT/US02/11524
- (22) International Filing Date: 12 April 2002 (12.04.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/283,948 16 April 2001 (16.04.2001) US  
60/284,443 18 April 2001 (18.04.2001) US
- (71) Applicant (for all designated States except US): **AMERICAN CYANAMID COMPANY** [US/US]; Five Giralda Farms, Madison, NJ 07940 (US).
- (72) Inventors; and  
(75) Inventors/Applicants (for US only): **ZAGURSKY, Robert, John** [US/US]; 569 Fox Hunt Drive, Victor, NY 14564 (US). **MASI, Amy, Wadhams** [US/US]; 326 Grand Circle, Caledonia, NY 14423 (US). **GREEN, Bruce, Arthur** [US/US]; 49 Northfield Gate, Pittsford, NY 14534 (US). **CHAKRAVARTI, Deb, Narayan** [IN/US]; 2 Fairway Crossing, Pittsford, NY 14534 (US). **RUSSELL, David, Parrish** [US/US]; 240 North Pleasant Street, Canandaigua, NY 14424 (US). **WOOTERS, Joseph,**
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**  
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 02/083855 A2

(54) Title: NOVEL STREPTOCOCCUS PNEUMONIAE OPEN READING FRAMES ENCODING POLYPEPTIDE ANTIGENS AND USES THEREOF

(57) Abstract: The present invention relates to newly identified open reading frames comprised within the genomic nucleotide sequence of *Streptococcus pneumoniae*, wherein the open reading frames encode polypeptides that are surface localized on *Streptococcus pneumoniae*. Thus, the invention relates to *Streptococcus pneumoniae* open reading frames that encode polypeptide antigens, polypeptides, preferably antigenic polypeptides, encoded by the *Streptococcus pneumoniae* open reading frames, vectors comprising open reading frame sequences and cells or animals transformed with these vectors. The invention relates also to methods of detecting these nucleic acids or polypeptides and kits for diagnosing *Streptococcus pneumoniae* infection. The invention finally relates to pharmaceutical compositions, in particular immunogenic compositions, for the prevention and/or treatment of bacterial infection, in particular infections with *Streptococcus pneumoniae*. In particular embodiments, the immunogenic compositions are used for the treatment or prevention of systemic diseases which are induced or exacerbated by *Streptococcus pneumoniae*. In other embodiments, the immunogenic compositions are used for the treatment or prevention of non-systemic diseases, particularly of the otitis media, which are induced or exacerbated by *Streptococcus pneumoniae*.

5                   **NOVEL *STREPTOCOCCUS PNEUMONIAE* OPEN READING FRAMES  
ENCODING POLYPEPTIDE ANTIGENS  
AND USES THEREOF**

10                   This application claims priority from copending provisional application serial  
number 60/283,948, filed on April 16, 2001, the entire disclosure of which is hereby  
incorporated by reference and provisional application serial number 60/284,443, filed  
April 18, 2001, the entire disclosure of which is hereby incorporated by reference.

**FIELD OF THE INVENTION**

15                   The invention relates to *Streptococcus pneumoniae* genomic sequence and  
polynucleotide sequences encoding polypeptides of *Streptococcus pneumoniae*.  
More particularly, the invention relates to newly identified polynucleotide open  
reading frames comprised within the genomic nucleotide sequence of *Streptococcus*  
*pneumoniae*, wherein the open reading frames encode *Streptococcus pneumoniae*  
20                   polypeptides, preferably polypeptides that are surface localized, secreted, membrane  
associated or exposed on *Streptococcus pneumoniae*.

**BACKGROUND OF THE INVENTION**

25                   *Streptococcus pneumoniae* infections are a major cause of human diseases  
such as otitis media, bacteremia, meningitis, septic arthritis and fatal pneumonia  
worldwide (Butler *et al.*, 1999; James and Thomas, 2000). Over the past 10-20  
years, *Streptococcus pneumoniae* has developed resistance to most antibiotics used  
for its treatment. In fact, it is common for *Streptococcus pneumoniae* to become  
resistant to more than one class of antibiotic, e.g.,  $\beta$ -lactams, macrolides,  
30                   lincosamides, trimethoprim-sulfamethoxazole, tetracyclines (Tauber, 2000), meaning  
*Streptococcus pneumoniae* treatment is becoming more difficult.

Thus, the rapid emergence of multi-drug resistant pneumococcal strains  
throughout the world has led to increased emphasis on prevention of pneumococcal

infections by immunization (Goldstein and Garau, 1997). The currently available 23-valent pneumococcal capsular polysaccharide vaccine, is not effective in children of less than 2 years of age or in immunocompromised patients, two of the major populations at risk from pneumococcal infection (Douglas *et al.*, 1983). A 7-valent pneumococcal polysaccharide-protein conjugate vaccine, recently licensed in the United States, was shown to be highly effective in infants and children against systemic pneumococcal disease caused by the vaccine serotypes and against cross-reactive capsular serotypes (Shinefield and Black, 2000). The seven capsular types cover greater than 80% of the invasive disease isolates in children in the United States, but only 57-60% of disease isolates in other areas of the world (Hausdorff *et al.*, 2000). There is therefore an immediate need for a cost-effective vaccine to cover most or all of the disease causing serotypes of pneumococci. While this can be achieved by adding conjugates covering additional serotypes, efforts continue to find non-capsular vaccine antigens that are conserved among all pneumococcal serotypes and effective against pneumococcal disease.

Protein antigens of *Streptococcus pneumoniae* have been evaluated for protective efficacy in animal models of pneumococcal infection. Some of the most commonly studied candidate antigens include the PspA proteins, PsaA lipoprotein, and the CbpA protein. Numerous studies have shown that PspA protein is a virulence factor (Crain *et al.*, 1990; McDaniel *et al.*, 1984) but it is antigenically variable among pneumococcal strains. A recent study has indicated that some antigenically conserved regions of a recombinant PspA variant may elicit cross-reactive antibodies in human adults (Nabors *et al.*, 2000). PsaA, a 37 kD lipoprotein with similarity to other gram-positive adhesins, is involved in  $Mn^{+}$  transport in pneumococci (Sampson *et al.*, 1994; Dintilhac *et al.*, 1997) and has also been shown to be protective in mouse models of systemic disease (Talkingington *et al.*, 1996). The surface exposed choline binding protein CbpA is antigenically conserved and protective in mouse models of pneumococcal disease (Rosenow *et al.*, 1997). Since nasopharyngeal colonization is a prerequisite for otic disease, intranasal immunization of mice with pneumococcal proteins and appropriate mucosal adjuvants has been used to enhance the mucosal antibody response and thus, the effectiveness of candidate antigens (Yamamoto *et al.*, 1998; Briles *et al.*, 2000).

While the PspA protein, PsaA lipoprotein and the CbpA protein antigens appear promising, it is possible that no one protein antigen will be effective against all *Streptococcus pneumoniae* serotypes. Laboratories therefore continue to search for additional candidates that are antigenically conserved and elicit antibodies that

5 reduce colonization (important for otitis media), are protective against systemic disease, or both. Thus, there is an immediate need for a cost-effective vaccine to cover most or all of the disease causing serotypes of *Streptococcus pneumoniae* and methods of diagnosing *Streptococcus pneumoniae* infection. A better understanding of the genetic and molecular levels of *Streptococcus pneumoniae* infection will

10 provide the basis for further development of preventative treatments, therapeutic treatments, new diagnostics and vaccine strategies which are specific for *Streptococcus pneumoniae*.

#### SUMMARY OF THE INVENTION

15 The present invention broadly relates to *Streptococcus pneumoniae* genomic sequence. More particularly, the invention relates to newly identified polynucleotide open reading frames comprised within the genomic nucleotide sequence of *Streptococcus pneumoniae*, wherein the open reading frames encode polypeptides that are surface localized, membrane associated, secreted, or exposed on

20 *Streptococcus pneumoniae*.

Thus, in certain aspects, the invention relates to *Streptococcus pneumoniae* open reading frames that encode *Streptococcus pneumoniae* polypeptides. In preferred embodiments, these *Streptococcus pneumoniae* polypeptides are antigenic polypeptides. As defined hereinafter, a *Streptococcus pneumoniae* antigenic

25 polypeptide, antigen or immunogen, is a *Streptococcus pneumoniae* polypeptide that is immunoreactive with an antibody or is a *Streptococcus pneumoniae* polypeptide that elicits an immune response. In other embodiments, the invention relates to the polynucleotides encoding these antigenic polypeptides. In other aspects, the invention relates to vectors comprising open reading frame sequences and cells or

30 animals transformed, transfected or infected with these vectors. The invention relates also to methods of detecting these nucleic acids or polypeptides and kits for diagnosing *Streptococcus pneumoniae* infection. The invention further relates to pharmaceutical compositions, in particular immunogenic compositions, for the

prevention and/or treatment of bacterial infection, in particular infections with *Streptococcus pneumoniae*. In a preferred embodiment, the immunogenic compositions are used for the treatment or prevention of systemic diseases that are induced or worsened by *Streptococcus pneumoniae*. In another preferred  
5 embodiment, the immunogenic compositions are used for the treatment or prevention of non-systemic diseases, particularly of the otitis media, which are induced or worsened by *Streptococcus pneumoniae*.

In particular embodiments, an isolated polynucleotide of the present invention is a polynucleotide comprising a nucleotide sequence having at least about 95%  
10 identity to a nucleotide sequence chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO:431 through SEQ ID NO:591, a degenerate variant thereof, or a fragment thereof. As defined hereinafter, a "degenerate variant" is defined as a polynucleotide that differs from the nucleotide sequence shown in SEQ ID NO:1 through SEQ ID NO:215 and SEQ ID NO:431 through SEQ ID NO:591 (and  
15 fragments thereof) due to degeneracy of the genetic code, but still encodes the same *Streptococcus pneumoniae* polypeptide (i.e., SEQ ID NO:216 through SEQ ID NO:430 and SEQ ID NO:592 through SEQ ID NO:752) as that encoded by the nucleotide sequence shown in SEQ ID NO:1 through SEQ ID NO:215 and SEQ ID NO:431 through SEQ ID NO:591.

20 In other embodiments, the polynucleotide is a complement to a nucleotide sequence chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO:431 through SEQ ID NO:591, a degenerate variant thereof, or a fragment thereof. In yet other embodiments, the polynucleotide is selected from the group consisting of DNA, chromosomal DNA, cDNA and RNA and may further comprise  
25 heterologous nucleotides.

In another embodiment, the invention comprises an isolated polynucleotide that hybridizes to a nucleotide sequence chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO:431 through SEQ ID NO:591, a complement thereof, a degenerate variant thereof, or a fragment thereof, under high stringency  
30 hybridization conditions. In yet other embodiments, the polynucleotide hybridizes under intermediate stringency hybridization conditions.

In a preferred embodiment, an isolated polynucleotide of a *Streptococcus pneumoniae* genomic sequence comprises a nucleotide sequence chosen from one

of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO:431 through SEQ ID NO:591, a fragment thereof, or a degenerate variant thereof, and encodes a polypeptide, a biological equivalent thereof, or a fragment thereof, selected from the group consisting of a *Streptococcus pneumoniae* polypeptide having 0, 1 or 2 transmembrane domains, a *Streptococcus pneumoniae* polypeptide having 3 or more transmembrane domains, a *Streptococcus pneumoniae* polypeptide having an outer membrane domain or a periplasmic domain, a *Streptococcus pneumoniae* polypeptide having an inner membrane domain, a *Streptococcus pneumoniae* polypeptide identified by Blastp analysis, a *Streptococcus pneumoniae* polypeptide identified by Pfam analysis, a *Streptococcus pneumoniae* lipoprotein, a *Streptococcus pneumoniae* polypeptide having a LPXTG motif, wherein the polypeptide is covalently attached to the peptidoglycan layer, a *Streptococcus pneumoniae* polypeptide having a peptidoglycan binding motif, wherein the polypeptide is associated with the peptidoglycan layer, a *Streptococcus pneumoniae* polypeptide having a signal sequence and a C-terminal Tyrosine or Phenylalanine amino acid, a *Streptococcus pneumoniae* polypeptide having a tripeptide RGD sequence, a *Streptococcus pneumoniae* polypeptide identified by proteomics as surface exposed and a *Streptococcus pneumoniae* polypeptide identified by proteomics as membrane associated.

In other embodiments, the isolated polynucleotide is a complement to a *Streptococcus pneumoniae* genomic sequence comprising a nucleotide sequence chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO:431 through SEQ ID NO:591, a fragment thereof, or a degenerate variant thereof, and encodes a polypeptide, a biological equivalent thereof, or a fragment thereof, selected from the group consisting of a *Streptococcus pneumoniae* polypeptide having 0, 1 or 2 transmembrane domains, a *Streptococcus pneumoniae* polypeptide having 3 or more transmembrane domains, a *Streptococcus pneumoniae* polypeptide having an outer membrane domain or a periplasmic domain, a *Streptococcus pneumoniae* polypeptide having an inner membrane domain, a *Streptococcus pneumoniae* polypeptide identified by Blastp analysis, a *Streptococcus pneumoniae* polypeptide identified by Pfam analysis, a *Streptococcus pneumoniae* lipoprotein, a *Streptococcus pneumoniae* polypeptide having a LPXTG motif, wherein the polypeptide is covalently attached to the peptidoglycan layer, a *Streptococcus*

*pneumoniae* polypeptide having a peptidoglycan binding motif, wherein the polypeptide is associated with the peptidoglycan layer, a *Streptococcus pneumoniae* polypeptide having a signal sequence and a C-terminal Tyrosine or Phenylalanine amino acid, a *Streptococcus pneumoniae* polypeptide having a tripeptide RGD sequence, a *Streptococcus pneumoniae* polypeptide identified by proteomics as surface exposed and a *Streptococcus pneumoniae* polypeptide identified by proteomics as membrane associated. In certain embodiments, the polynucleotide is selected from the group consisting of DNA, chromosomal DNA, cDNA and RNA and may further comprise heterologous nucleotides. In still other embodiments, the polynucleotide encodes a fusion polypeptide.

In a preferred embodiment, a polynucleotide encoding a polypeptide having 0, 1 or 2 transmembrane domains comprises a nucleotide sequence chosen from one of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 39, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 45, SEQ ID NO: 47, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 53, SEQ ID NO: 55, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70, SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 78, SEQ ID NO: 79, SEQ ID NO: 81, SEQ ID NO: 83, SEQ ID NO: 85, SEQ ID NO: 86, SEQ ID NO: 89, SEQ ID NO: 91, SEQ ID NO: 92, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 97, SEQ ID NO: 100, SEQ ID NO: 104, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 109, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 113, SEQ ID NO: 116, SEQ ID NO: 121, SEQ ID NO: 122, SEQ ID NO: 123, SEQ ID NO: 125, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 128, SEQ ID NO: 131, SEQ ID NO: 132, SEQ ID NO: 134, SEQ ID NO: 136, SEQ ID NO: 137, SEQ ID NO: 138, SEQ ID NO: 141, SEQ ID NO: 142, SEQ ID NO: 143, SEQ ID NO: 144, SEQ ID NO: 147, SEQ ID NO: 148, SEQ ID NO: 149, SEQ ID NO: 150, SEQ ID NO: 155, SEQ ID NO: 156, SEQ ID NO: 158, SEQ ID NO: 161, SEQ ID NO: 162, SEQ ID NO: 165, SEQ ID NO: 170, SEQ ID NO: 171, SEQ ID NO: 172, SEQ ID NO: 174, SEQ ID NO: 176, SEQ ID NO: 179, SEQ ID NO: 183, SEQ ID NO: 185, SEQ ID

NO: 187, SEQ ID NO: 192, SEQ ID NO: 195, SEQ ID NO: 196, SEQ ID NO: 197, SEQ ID NO: 199, SEQ ID NO: 200, SEQ ID NO: 201, SEQ ID NO: 202, SEQ ID NO: 204, SEQ ID NO: 205, SEQ ID NO: 207, SEQ ID NO: 209 and SEQ ID NO: 210.

- In another preferred embodiment, a polynucleotide encoding a polypeptide
- 5 having 3 or more transmembrane domains comprises a nucleotide sequence chosen from one of SEQ ID NO: 2, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 43, SEQ ID
- 10 NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, SEQ ID NO: 52, SEQ ID NO: 54, SEQ ID NO: 56, SEQ ID NO: 59, SEQ ID NO: 65, SEQ ID NO: 71, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 80, SEQ ID NO: 82, SEQ ID NO: 84, SEQ ID NO: 87, SEQ ID NO: 88, SEQ ID NO: 90, SEQ ID NO: 93, SEQ ID NO: 94, SEQ ID NO: 98, SEQ ID NO: 99, SEQ ID NO: 101, SEQ ID NO: 102, SEQ ID NO: 103, SEQ
- 15 ID NO: 107, SEQ ID NO: 108, SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 117, SEQ ID NO: 118, SEQ ID NO: 119, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 129, SEQ ID NO: 130, SEQ ID NO: 133, SEQ ID NO: 135, SEQ ID NO: 139, SEQ ID NO: 140, SEQ ID NO: 145, SEQ ID NO: 146, SEQ ID NO: 151, SEQ ID NO: 152, SEQ ID NO: 153, SEQ ID NO: 154, SEQ ID NO: 157, SEQ ID NO:
- 20 159, SEQ ID NO: 160, SEQ ID NO: 163, SEQ ID NO: 164, SEQ ID NO: 166, SEQ ID NO: 167, SEQ ID NO: 168, SEQ ID NO: 169, SEQ ID NO: 173, SEQ ID NO: 175, SEQ ID NO: 177, SEQ ID NO: 178, SEQ ID NO: 180, SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 184, SEQ ID NO: 186, SEQ ID NO: 188, SEQ ID NO: 189, SEQ ID NO: 190, SEQ ID NO: 191, SEQ ID NO: 193, SEQ ID NO: 194, SEQ ID NO: 198,
- 25 SEQ ID NO: 203, SEQ ID NO: 206, SEQ ID NO: 208, SEQ ID NO: 211, SEQ ID NO: 212, SEQ ID NO: 213, SEQ ID NO: 214 and SEQ ID NO: 215.

- In other preferred embodiments, a polynucleotide encoding a polypeptide having an outer membrane domain or a periplasmic domain comprises a nucleotide sequence chosen from one of SEQ ID NO: 3, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID
- 30 NO: 23, SEQ ID NO: 39, SEQ ID NO: 50, SEQ ID NO: 62, SEQ ID NO: 67, SEQ ID NO: 78, SEQ ID NO: 85, SEQ ID NO: 125, SEQ ID NO: 134, SEQ ID NO: 147, SEQ ID NO: 165, SEQ ID NO: 172 and SEQ ID NO: 179.



In other preferred embodiments, a polynucleotide encoding a polypeptide having an inner membrane domain comprises a nucleotide sequence chosen from one of SEQ ID NO: 2, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 56, SEQ ID NO: 59, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 65, SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70, SEQ ID NO: 71, SEQ ID NO: 73, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 81, SEQ ID NO: 82, SEQ ID NO: 83, SEQ ID NO: 84, SEQ ID NO: 86, SEQ ID NO: 87, SEQ ID NO: 88, SEQ ID NO: 90, SEQ ID NO: 91, SEQ ID NO: 93, SEQ ID NO: 94, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 97, SEQ ID NO: 98, SEQ ID NO: 99, SEQ ID NO: 100, SEQ ID NO: 101, SEQ ID NO: 102, SEQ ID NO: 103, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 107, SEQ ID NO: 108, SEQ ID NO: 109, SEQ ID NO: 112, SEQ ID NO: 113, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 117, SEQ ID NO: 118, SEQ ID NO: 119, SEQ ID NO: 120, SEQ ID NO: 121, SEQ ID NO: 122, SEQ ID NO: 123, SEQ ID NO: 124, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 130, SEQ ID NO: 131, SEQ ID NO: 132, SEQ ID NO: 133, SEQ ID NO: 135, SEQ ID NO: 136, SEQ ID NO: 139, SEQ ID NO: 140, SEQ ID NO: 141, SEQ ID NO: 142, SEQ ID NO: 144, SEQ ID NO: 145, SEQ ID NO: 146, SEQ ID NO: 148, SEQ ID NO: 150, SEQ ID NO: 151, SEQ ID NO: 152, SEQ ID NO: 153, SEQ ID NO: 154, SEQ ID NO: 156, SEQ ID NO: 157, SEQ ID NO: 158, SEQ ID NO: 159, SEQ ID NO: 160, SEQ ID NO: 162, SEQ ID NO: 163, SEQ ID NO: 164, SEQ ID NO: 166, SEQ ID NO: 167, SEQ ID NO: 168, SEQ ID NO: 169, SEQ ID NO: 170, SEQ ID NO: 173, SEQ ID NO: 175, SEQ ID NO: 176, SEQ ID NO: 177, SEQ ID NO: 178, SEQ ID NO: 180, SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 184, SEQ ID NO: 186, SEQ ID NO: 187, SEQ ID NO: 188, SEQ ID NO: 189, SEQ ID NO: 190, SEQ ID NO: 191, SEQ ID NO: 192, SEQ ID NO: 193, SEQ ID NO: 194, SEQ ID NO: 195, SEQ ID NO: 198, SEQ ID NO: 200, SEQ ID NO: 203, SEQ ID

NO: 206, SEQ ID NO: 208, SEQ ID NO: 209, SEQ ID NO: 211, SEQ ID NO: 212, SEQ ID NO: 213, SEQ ID NO: 214 and SEQ ID NO: 215.

In yet another preferred embodiment, a polynucleotide encoding a polypeptide identified by Blastp analysis comprises a nucleotide sequence chosen from one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 7, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 16, SEQ ID NO: 20, SEQ ID NO: 24, SEQ ID NO: 27, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 48, SEQ ID NO: 51, SEQ ID NO: 53, SEQ ID NO: 59, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 65, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70, SEQ ID NO: 71, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 87, SEQ ID NO: 88, SEQ ID NO: 90, SEQ ID NO: 94, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 98, SEQ ID NO: 100, SEQ ID NO: 103, SEQ ID NO: 105, SEQ ID NO: 107, SEQ ID NO: 108, SEQ ID NO: 109, SEQ ID NO: 112, SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 117, SEQ ID NO: 118, SEQ ID NO: 122, SEQ ID NO: 123, SEQ ID NO: 124, SEQ ID NO: 127, SEQ ID NO: 129, SEQ ID NO: 131, SEQ ID NO: 132, SEQ ID NO: 133, SEQ ID NO: 134, SEQ ID NO: 135, SEQ ID NO: 136, SEQ ID NO: 138, SEQ ID NO: 139, SEQ ID NO: 141, SEQ ID NO: 144, SEQ ID NO: 146, SEQ ID NO: 147, SEQ ID NO: 151, SEQ ID NO: 152, SEQ ID NO: 154, SEQ ID NO: 155, SEQ ID NO: 157, SEQ ID NO: 158, SEQ ID NO: 159, SEQ ID NO: 160, SEQ ID NO: 161, SEQ ID NO: 162, SEQ ID NO: 163, SEQ ID NO: 165, SEQ ID NO: 166, SEQ ID NO: 167, SEQ ID NO: 169, SEQ ID NO: 172, SEQ ID NO: 173, SEQ ID NO: 176, SEQ ID NO: 177, SEQ ID NO: 178, SEQ ID NO: 180, SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 184, SEQ ID NO: 185, SEQ ID NO: 186, SEQ ID NO: 188, SEQ ID NO: 189, SEQ ID NO: 191, SEQ ID NO: 193, SEQ ID NO: 196, SEQ ID NO: 197, SEQ ID NO: 198, SEQ ID NO: 199, SEQ ID NO: 200, SEQ ID NO: 201, SEQ ID NO: 202, SEQ ID NO: 204, SEQ ID NO: 205, SEQ ID NO: 206, SEQ ID NO: 207, SEQ ID NO: 208, SEQ ID NO: 210, SEQ ID NO: 212, SEQ ID NO: 213 and SEQ ID NO: 214.

In still further preferred embodiments, a polynucleotide encoding a polypeptide identified by Pfam analysis comprises a nucleotide sequence chosen from one of SEQ ID NO: 4, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 41, SEQ ID NO: 45, SEQ ID NO: 55, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 63, SEQ ID

- NO: 64, SEQ ID NO: 66, SEQ ID NO: 72, SEQ ID NO: 74, SEQ ID NO: 89, SEQ ID NO: 92, SEQ ID NO: 104, SEQ ID NO: 111, SEQ ID NO: 116, SEQ ID NO: 119, SEQ ID NO: 128, SEQ ID NO: 137, SEQ ID NO: 142, SEQ ID NO: 143, SEQ ID NO: 149, SEQ ID NO: 151, SEQ ID NO: 152, SEQ ID NO: 153, SEQ ID NO: 157, SEQ ID NO: 159, SEQ ID NO: 160, SEQ ID NO: 162, SEQ ID NO: 163, SEQ ID NO: 164, SEQ ID NO: 165, SEQ ID NO: 166, SEQ ID NO: 169, SEQ ID NO: 171, SEQ ID NO: 174, SEQ ID NO: 176, SEQ ID NO: 180, SEQ ID NO: 182, SEQ ID NO: 183, SEQ ID NO: 184, SEQ ID NO: 186, SEQ ID NO: 188, SEQ ID NO: 189, SEQ ID NO: 195, SEQ ID NO: 198, SEQ ID NO: 199, SEQ ID NO: 205, SEQ ID NO: 212 and SEQ ID NO: 213.
- 10 In another preferred embodiment, a polynucleotide encoding a lipoprotein comprises a nucleotide sequence chosen from one of SEQ ID NO: 3, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 13, SEQ ID NO: 21, SEQ ID NO: 26, SEQ ID NO: 34, SEQ ID NO: 62, SEQ ID NO: 67, SEQ ID NO: 85, SEQ ID NO: 134, SEQ ID NO: 147, SEQ ID NO: 150, SEQ ID NO: 168, SEQ ID NO: 170 and SEQ ID NO: 173.
- 15 In other preferred embodiments, a polynucleotide encoding a polypeptide having a LPXTG motif and is covalently attached to the peptidoglycan layer comprises a nucleotide sequence chosen from one of SEQ ID NO: 13, SEQ ID NO: 21, SEQ ID NO: 34 and SEQ ID NO: 170; or a polynucleotide encoding a polypeptide having a peptidoglycan binding motif and associated with the peptidoglycan layer
- 20 comprises a nucleotide sequence chosen from one of SEQ ID NO: 25, SEQ ID NO: 49 and SEQ ID NO: 110.
- In another preferred embodiment, a polynucleotide encoding a polypeptide having a signal sequence and a C-terminal Tyrosine or Phenylalanine amino acid comprises a nucleotide sequence chosen from one of SEQ ID NO:11, SEQ ID NO:39, SEQ ID NO:73, SEQ ID NO:97, SEQ ID NO:106, SEQ ID NO: 125 and SEQ ID NO:187.
- 25 In yet another preferred embodiment, a polynucleotide encoding a polypeptide having a tripeptide RGD sequence that potentially is involved in cell attachment comprises a nucleotide sequence chosen from one of SEQ ID NO:1, SEQ ID NO:21, SEQ ID NO:66 and SEQ ID NO:67.
- 30 In another preferred embodiment, a polynucleotide encoding a polypeptide identified by proteomics as surface exposed comprises a nucleotide sequence chosen from one of SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:46,

SEQ ID NO:64, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:69, SEQ ID NO:71, SEQ ID NO:74, SEQ ID NO:91, SEQ ID NO:103, SEQ ID NO:116, SEQ ID NO:128, SEQ ID NO:131, SEQ ID NO:136, SEQ ID NO:151, SEQ ID NO:156, SEQ ID NO:159, SEQ ID NO:162, SEQ ID NO:164, SEQ ID NO:172, SEQ ID NO:176, SEQ ID NO:178, SEQ ID NO:179, SEQ ID NO:180, SEQ ID NO:182 and SEQ ID NO:205.

In still another embodiment, a polynucleotide encoding a polypeptide identified by proteomics as membrane associated comprises a nucleotide sequence chosen from one of SEQ ID NO:431 through SEQ ID NO:591.

In certain aspects, the invention relates to *Streptococcus pneumoniae* polypeptides. More particularly, the invention relates to *Streptococcus pneumoniae* polypeptides, more preferably antigenic polypeptides, encoded by *Streptococcus pneumoniae* polynucleotide open reading frames. Thus, in certain embodiments, an isolated polypeptide is encoded by a polynucleotide comprising a nucleotide sequence having at least about 95% identity to a nucleotide sequence chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO: 431 through SEQ ID NO: 591, a degenerate variant thereof, or a fragment thereof. In a preferred embodiment, the isolated polypeptide encoded by one of the above polynucleotides comprises an amino acid sequence having at least about 95% identity to an amino acid sequence chosen from one of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID NO: 592 through SEQ ID NO: 752, a biological equivalent thereof, or a fragment thereof. In other embodiments, the polypeptide is a fusion polypeptide. In a preferred embodiment, the polypeptide immunoreacts with seropositive serum of an individual infected with *Streptococcus pneumoniae*.

In preferred embodiments, the isolated polypeptide encoded by a polynucleotide comprising a nucleotide sequence having at least about 95% identity to a nucleotide sequence chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO: 431 through SEQ ID NO: 591, a degenerate variant thereof, or a fragment thereof, is further defined as a *Streptococcus pneumoniae* polypeptide having 0, 1 or 2 transmembrane domains, a *Streptococcus pneumoniae* polypeptide having 3 or more transmembrane domains, a *Streptococcus pneumoniae* polypeptide having an outer membrane domain or a periplasmic domain, a *Streptococcus pneumoniae* polypeptide having an inner membrane domain, a *Streptococcus pneumoniae* polypeptide identified by Blastp analysis, a *Streptococcus pneumoniae*

polypeptide identified by Pfam analysis, a *Streptococcus pneumoniae* lipoprotein, a *Streptococcus pneumoniae* polypeptide having a LPXTG motif, wherein the polypeptide is covalently attached to the peptidoglycan layer, a *Streptococcus pneumoniae* polypeptide having a peptidoglycan binding motif, wherein the

5 polypeptide is associated with the peptidoglycan layer, a *Streptococcus pneumoniae* polypeptide having a signal sequence and a C-terminal Tyrosine or Phenylalanine amino acid, a *Streptococcus pneumoniae* polypeptide having a tripeptide RGD sequence, a *Streptococcus pneumoniae* polypeptide identified by proteomics as surface exposed or a *Streptococcus pneumoniae* polypeptide identified by

10 proteomics as membrane associated, where each of these groups has the set of ORFs identified above as within SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO: 431 through SEQ ID NO: 591.

In a particularly preferred embodiment, an isolated polypeptide comprises an amino acid sequence having at least about 95% identity to an amino acid sequence

15 chosen from one of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID NO: 592 through SEQ ID NO: 752, a biological equivalent thereof, or a fragment thereof. In another embodiment, the polypeptide is a fusion polypeptide. In a particularly preferred embodiment, the polypeptide immunoreacts with seropositive serum of an individual infected with *Streptococcus pneumoniae*. In yet other preferred

20 embodiments, the polypeptide is further defined as a *Streptococcus pneumoniae* polypeptide having 0, 1 or 2 transmembrane domains, a *Streptococcus pneumoniae* polypeptide having 3 or more transmembrane domains, a *Streptococcus pneumoniae* polypeptide having an outer membrane domain or a periplasmic domain, a *Streptococcus pneumoniae* polypeptide having an inner membrane domain, a

25 *Streptococcus pneumoniae* polypeptide identified by Blastp analysis, a *Streptococcus pneumoniae* polypeptide identified by Pfam analysis, a *Streptococcus pneumoniae* lipoprotein, a *Streptococcus pneumoniae* polypeptide having a LPXTG motif, wherein the polypeptide is covalently attached to the peptidoglycan layer, a *Streptococcus pneumoniae* polypeptide having a peptidoglycan binding motif,

30 wherein the polypeptide is associated with the peptidoglycan layer, a *Streptococcus pneumoniae* polypeptide having a signal sequence and a C-terminal Tyrosine or Phenylalanine amino acid, a *Streptococcus pneumoniae* polypeptide having a tripeptide RGD sequence, a *Streptococcus pneumoniae* polypeptide identified by

proteomics as surface exposed or a *Streptococcus pneumoniae* polypeptide identified by proteomics as membrane associated.

In a preferred embodiment, a polypeptide having 0, 1 or 2 transmembrane domains comprises an amino acid sequence chosen from one of SEQ ID NO: 216, 5 SEQ ID NO: 218, SEQ ID NO: 219, SEQ ID NO: 222, SEQ ID NO: 223, SEQ ID NO: 224, SEQ ID NO: 226, SEQ ID NO: 228, SEQ ID NO: 231, SEQ ID NO: 232, SEQ ID NO: 233, SEQ ID NO: 234, SEQ ID NO: 237, SEQ ID NO: 238, SEQ ID NO: 239, SEQ ID NO: 240, SEQ ID NO: 243, SEQ ID NO: 244, SEQ ID NO: 247, SEQ ID NO: 249, SEQ ID NO: 251, SEQ ID NO: 254, SEQ ID NO: 256, SEQ ID NO: 257, SEQ ID NO: 260, SEQ ID NO: 262, SEQ ID NO: 264, SEQ ID NO: 265, SEQ ID NO: 266, 10 SEQ ID NO: 268, SEQ ID NO: 270, SEQ ID NO: 272, SEQ ID NO: 273, SEQ ID NO: 275, SEQ ID NO: 276, SEQ ID NO: 277, SEQ ID NO: 278, SEQ ID NO: 279, SEQ ID NO: 281, SEQ ID NO: 282, SEQ ID NO: 283, SEQ ID NO: 284, SEQ ID NO: 285, SEQ ID NO: 286, SEQ ID NO: 287, SEQ ID NO: 289, SEQ ID NO: 293, SEQ ID NO: 294, SEQ ID NO: 296, SEQ ID NO: 298, SEQ ID NO: 300, SEQ ID NO: 301, SEQ ID NO: 304, SEQ ID NO: 306, SEQ ID NO: 307, SEQ ID NO: 310, SEQ ID NO: 311, SEQ ID NO: 312, SEQ ID NO: 315, SEQ ID NO: 319, SEQ ID NO: 320, SEQ ID NO: 321, SEQ ID NO: 324, SEQ ID NO: 325, SEQ ID NO: 326, SEQ ID NO: 328, SEQ ID NO: 331, SEQ ID NO: 336, SEQ ID NO: 337, SEQ ID NO: 338, SEQ ID NO: 340, 20 SEQ ID NO: 341, SEQ ID NO: 342, SEQ ID NO: 343, SEQ ID NO: 346, SEQ ID NO: 347, SEQ ID NO: 349, SEQ ID NO: 351, SEQ ID NO: 352, SEQ ID NO: 353, SEQ ID NO: 356, SEQ ID NO: 357, SEQ ID NO: 358, SEQ ID NO: 359, SEQ ID NO: 362, SEQ ID NO: 363, SEQ ID NO: 364, SEQ ID NO: 365, SEQ ID NO: 370, SEQ ID NO: 371, SEQ ID NO: 373, SEQ ID NO: 376, SEQ ID NO: 377, SEQ ID NO: 380, SEQ ID NO: 385, SEQ ID NO: 386, SEQ ID NO: 387, SEQ ID NO: 389, SEQ ID NO: 391, 25 SEQ ID NO: 394, SEQ ID NO: 398, SEQ ID NO: 400, SEQ ID NO: 402, SEQ ID NO: 407, SEQ ID NO: 410, SEQ ID NO: 411, SEQ ID NO: 412, SEQ ID NO: 414, SEQ ID NO: 415, SEQ ID NO: 416, SEQ ID NO: 417, SEQ ID NO: 419, SEQ ID NO: 420, SEQ ID NO: 422, SEQ ID NO: 424, SEQ ID NO: 425, a biological equivalent thereof, 30 or a fragment thereof.

In another preferred embodiment, a polypeptide having 3 or more transmembrane domains comprises an amino acid sequence chosen from one of SEQ ID NO: 217, SEQ ID NO: 220, SEQ ID NO: 221, SEQ ID NO: 225, SEQ ID NO:

227, SEQ ID NO: 229, SEQ ID NO: 230, SEQ ID NO: 235, SEQ ID NO: 236, SEQ ID NO: 241, SEQ ID NO: 242, SEQ ID NO: 245, SEQ ID NO: 246, SEQ ID NO: 248, SEQ ID NO: 250, SEQ ID NO: 252, SEQ ID NO: 253, SEQ ID NO: 255, SEQ ID NO: 258, SEQ ID NO: 259, SEQ ID NO: 261, SEQ ID NO: 263, SEQ ID NO: 267, SEQ ID NO: 269, SEQ ID NO: 271, SEQ ID NO: 274, SEQ ID NO: 280, SEQ ID NO: 286, SEQ ID NO: 290, SEQ ID NO: 291, SEQ ID NO: 292, SEQ ID NO: 295, SEQ ID NO: 297, SEQ ID NO: 299, SEQ ID NO: 302, SEQ ID NO: 303, SEQ ID NO: 305, SEQ ID NO: 308, SEQ ID NO: 309, SEQ ID NO: 313, SEQ ID NO: 314, SEQ ID NO: 316, SEQ ID NO: 317, SEQ ID NO: 318, SEQ ID NO: 322, SEQ ID NO: 323, SEQ ID NO: 327, SEQ ID NO: 329, SEQ ID NO: 330, SEQ ID NO: 332, SEQ ID NO: 333, SEQ ID NO: 334, SEQ ID NO: 335, SEQ ID NO: 339, SEQ ID NO: 344, SEQ ID NO: 345, SEQ ID NO: 348, SEQ ID NO: 350, SEQ ID NO: 354, SEQ ID NO: 355, SEQ ID NO: 360, SEQ ID NO: 361, SEQ ID NO: 366, SEQ ID NO: 367, SEQ ID NO: 368, SEQ ID NO: 369, SEQ ID NO: 372, SEQ ID NO: 374, SEQ ID NO: 375, SEQ ID NO: 378, SEQ ID NO: 379, SEQ ID NO: 381, SEQ ID NO: 382, SEQ ID NO: 383, SEQ ID NO: 384, SEQ ID NO: 388, SEQ ID NO: 390, SEQ ID NO: 392, SEQ ID NO: 393, SEQ ID NO: 395, SEQ ID NO: 396, SEQ ID NO: 397, SEQ ID NO: 399, SEQ ID NO: 401, SEQ ID NO: 403, SEQ ID NO: 404, SEQ ID NO: 405, SEQ ID NO: 406, SEQ ID NO: 408, SEQ ID NO: 409, SEQ ID NO: 413, SEQ ID NO: 418, SEQ ID NO: 421, SEQ ID NO: 423, SEQ ID NO: 426, SEQ ID NO: 427, SEQ ID NO: 428, SEQ ID NO: 429, SEQ ID NO: 430, a biological equivalent thereof, or a fragment thereof.

In yet other preferred embodiments, a polypeptide having an outer membrane domain or a periplasmic domain comprises an amino acid sequence chosen from one of SEQ ID NO: 218, SEQ ID NO: 223, SEQ ID NO: 224, SEQ ID NO: 238, SEQ ID NO: 254, SEQ ID NO: 265, SEQ ID NO: 277, SEQ ID NO: 282, SEQ ID NO: 293, SEQ ID NO: 300, SEQ ID NO: 340, SEQ ID NO: 349, SEQ ID NO: 362, SEQ ID NO: 380, SEQ ID NO: 387, SEQ ID NO: 394, a biological equivalent thereof, or a fragment thereof.

In yet other preferred embodiments, a polynucleotide encoding a polypeptide having an inner membrane domain comprises an amino acid sequence chosen from one of SEQ ID NO: 217, SEQ ID NO: 220, SEQ ID NO: 221, SEQ ID NO: 222, SEQ ID NO: 225, SEQ ID NO: 226, SEQ ID NO: 227, SEQ ID NO: 228, SEQ ID NO: 229, SEQ ID NO: 230, SEQ ID NO: 231, SEQ ID NO: 232, SEQ ID NO: 234, SEQ ID NO:

235, SEQ ID NO: 236, SEQ ID NO: 237, SEQ ID NO: 241, SEQ ID NO: 242, SEQ ID  
NO: 243, SEQ ID NO: 244, SEQ ID NO: 245, SEQ ID NO: 246, SEQ ID NO: 247,  
SEQ ID NO: 248, SEQ ID NO: 249, SEQ ID NO: 250, SEQ ID NO: 251, SEQ ID NO:  
252, SEQ ID NO: 253, SEQ ID NO: 255, SEQ ID NO: 258, SEQ ID NO: 259, SEQ ID  
5 NO: 261, SEQ ID NO: 262, SEQ ID NO: 263, SEQ ID NO: 266, SEQ ID NO: 267,  
SEQ ID NO: 268, SEQ ID NO: 269, SEQ ID NO: 271, SEQ ID NO: 274, SEQ ID NO:  
275, SEQ ID NO: 276, SEQ ID NO: 280, SEQ ID NO: 283, SEQ ID NO: 284, SEQ ID  
NO: 285, SEQ ID NO: 286, SEQ ID NO: 288, SEQ ID NO: 290, SEQ ID NO: 291,  
SEQ ID NO: 292, SEQ ID NO: 294, SEQ ID NO: 295, SEQ ID NO: 296, SEQ ID NO:  
10 297, SEQ ID NO: 298, SEQ ID NO: 299, SEQ ID NO: 301 SEQ ID NO: 302, SEQ ID  
NO: 303, SEQ ID NO: 305, SEQ ID NO: 306, SEQ ID NO: 308, SEQ ID NO: 309,  
SEQ ID NO: 310, SEQ ID NO: 311, SEQ ID NO: 312, SEQ ID NO: 313, SEQ ID NO:  
314, SEQ ID NO: 315, SEQ ID NO: 316, SEQ ID NO: 317, SEQ ID NO: 318, SEQ ID  
NO: 320, SEQ ID NO: 321, SEQ ID NO: 322, SEQ ID NO: 323, SEQ ID NO: 324,  
15 SEQ ID NO: 327, SEQ ID NO: 328, SEQ ID NO: 329, SEQ ID NO: 330, SEQ ID NO:  
332, SEQ ID NO: 333, SEQ ID NO: 334, SEQ ID NO: 335, SEQ ID NO: 336, SEQ ID  
NO: 337, SEQ ID NO: 338, SEQ ID NO: 339, SEQ ID NO: 341, SEQ ID NO: 342,  
SEQ ID NO: 343, SEQ ID NO: 344, SEQ ID NO: 345, SEQ ID NO: 346, SEQ ID NO:  
347, SEQ ID NO: 348, SEQ ID NO: 350, SEQ ID NO: 351, SEQ ID NO: 354, SEQ ID  
20 NO: 355, SEQ ID NO: 356, SEQ ID NO: 357, SEQ ID NO: 359, SEQ ID NO: 360,  
SEQ ID NO: 361, SEQ ID NO: 362, SEQ ID NO: 365, SEQ ID NO: 366, SEQ ID NO:  
367, SEQ ID NO: 368, SEQ ID NO: 369, SEQ ID NO: 371, SEQ ID NO: 372, SEQ ID  
NO: 373, SEQ ID NO: 374, SEQ ID NO: 375, SEQ ID NO: 377, SEQ ID NO: 378,  
SEQ ID NO: 379, SEQ ID NO: 381, SEQ ID NO: 382, SEQ ID NO: 383, SEQ ID NO:  
25 384, SEQ ID NO: 385, SEQ ID NO: 388, SEQ ID NO: 390, SEQ ID NO: 391, SEQ ID  
NO: 392, SEQ ID NO: 393, SEQ ID NO: 395, SEQ ID NO: 396, SEQ ID NO: 397,  
SEQ ID NO: 399, SEQ ID NO: 401, SEQ ID NO: 402, SEQ ID NO: 403, SEQ ID NO:  
404, SEQ ID NO: 405, SEQ ID NO: 406, SEQ ID NO: 407, SEQ ID NO: 408, SEQ ID  
NO: 409, SEQ ID NO: 410, SEQ ID NO: 413, SEQ ID NO: 415, SEQ ID NO: 418,  
30 SEQ ID NO: 421, SEQ ID NO: 423, SEQ ID NO: 424, SEQ ID NO: 426, SEQ ID NO:  
427, SEQ ID NO: 428, SEQ ID NO: 429, SEQ ID NO: 430, a biological equivalent  
thereof, or a fragment thereof.



In still another preferred embodiment, a polypeptide identified by Blastp analysis comprises an amino acid sequence chosen from one of SEQ ID NO: 216, SEQ ID NO: 217, SEQ ID NO: 222, SEQ ID NO: 225, SEQ ID NO: 227, SEQ ID NO: 231, SEQ ID NO: 235, SEQ ID NO: 239, SEQ ID NO: 242, SEQ ID NO: 245, SEQ ID NO: 246, SEQ ID NO: 247, SEQ ID NO: 248, SEQ ID NO: 249, SEQ ID NO: 250, SEQ ID NO: 253, SEQ ID NO: 255, SEQ ID NO: 257, SEQ ID NO: 258, SEQ ID NO: 259, SEQ ID NO: 263, SEQ ID NO: 266, SEQ ID NO: 268, SEQ ID NO: 269, SEQ ID NO: 275, SEQ ID NO: 276, SEQ ID NO: 280, SEQ ID NO: 282, SEQ ID NO: 283, SEQ ID NO: 284, SEQ ID NO: 285, SEQ ID NO: 286, SEQ ID NO: 290, SEQ ID NO: 291, SEQ ID NO: 292, SEQ ID NO: 293, SEQ ID NO: 294, SEQ ID NO: 295, SEQ ID NO: 302, SEQ ID NO: 303, SEQ ID NO: 305, SEQ ID NO: 309, SEQ ID NO: 310, SEQ ID NO: 311, SEQ ID NO: 313, SEQ ID NO: 315, SEQ ID NO: 318, SEQ ID NO: 320, SEQ ID NO: 322, SEQ ID NO: 323, SEQ ID NO: 324, SEQ ID NO: 327, SEQ ID NO: 328, SEQ ID NO: 330, SEQ ID NO: 332, SEQ ID NO: 333, SEQ ID NO: 337, SEQ ID NO: 338, SEQ ID NO: 339, SEQ ID NO: 342, SEQ ID NO: 344, SEQ ID NO: 346, SEQ ID NO: 347, SEQ ID NO: 348, SEQ ID NO: 349, SEQ ID NO: 350, SEQ ID NO: 351, SEQ ID NO: 353, SEQ ID NO: 354, SEQ ID NO: 356, SEQ ID NO: 359, SEQ ID NO: 361, SEQ ID NO: 362, SEQ ID NO: 366, SEQ ID NO: 367, SEQ ID NO: 369, SEQ ID NO: 370, SEQ ID NO: 372, SEQ ID NO: 373, SEQ ID NO: 374, SEQ ID NO: 375, SEQ ID NO: 376, SEQ ID NO: 377, SEQ ID NO: 378, SEQ ID NO: 380, SEQ ID NO: 381, SEQ ID NO: 382, SEQ ID NO: 384, SEQ ID NO: 387, SEQ ID NO: 388, SEQ ID NO: 391, SEQ ID NO: 392, SEQ ID NO: 393, SEQ ID NO: 395, SEQ ID NO: 396, SEQ ID NO: 397, SEQ ID NO: 399, SEQ ID NO: 400, SEQ ID NO: 401, SEQ ID NO: 403, SEQ ID NO: 404, SEQ ID NO: 406, SEQ ID NO: 408, SEQ ID NO: 411, SEQ ID NO: 412, SEQ ID NO: 413, SEQ ID NO: 414, SEQ ID NO: 415, SEQ ID NO: 416, SEQ ID NO: 417, SEQ ID NO: 419, SEQ ID NO: 420, SEQ ID NO: 421, SEQ ID NO: 422, SEQ ID NO: 423, SEQ ID NO: 425, SEQ ID NO: 427, SEQ ID NO: 428, SEQ ID NO: 429, a biological equivalent thereof, or a fragment thereof.

In other preferred embodiments, a polypeptide identified by Pfam analysis comprises an amino acid sequence chosen from one of SEQ ID NO: 219, SEQ ID NO: 233, SEQ ID NO: 234, SEQ ID NO: 255, SEQ ID NO: 260, SEQ ID NO: 270, SEQ ID NO: 272, SEQ ID NO: 273, SEQ ID NO: 278, SEQ ID NO: 279, SEQ ID NO: 281, SEQ ID NO: 287, SEQ ID NO: 289, SEQ ID NO: 304, SEQ ID NO: 307, SEQ ID

- NO: 319, SEQ ID NO: 326, SEQ ID NO: 331, SEQ ID NO: 334, SEQ ID NO: 343, SEQ ID NO: 352, SEQ ID NO: 357, SEQ ID NO: 358, SEQ ID NO: 364, SEQ ID NO: 366, SEQ ID NO: 367, SEQ ID NO: 368, SEQ ID NO: 372, SEQ ID NO: 374, SEQ ID NO: 375, SEQ ID NO: 377, SEQ ID NO: 378, SEQ ID NO: 379, SEQ ID NO: 380, 5 SEQ ID NO: 381, SEQ ID NO: 384, SEQ ID NO: 386, SEQ ID NO: 389, SEQ ID NO: 391, SEQ ID NO: 395, SEQ ID NO: 397, SEQ ID NO: 398, SEQ ID NO: 399, SEQ ID NO: 401, SEQ ID NO: 403, SEQ ID NO: 404, SEQ ID NO: 410, SEQ ID NO: 413, SEQ ID NO: 414, SEQ ID NO: 420, SEQ ID NO: 427, SEQ ID NO: 428, a biological equivalent thereof, or a fragment thereof.
- 10 In one preferred embodiment, a polypeptide is a lipoprotein and comprises an amino acid sequence chosen from one of SEQ ID NO: 218, SEQ ID NO: 223, SEQ ID NO: 224, SEQ ID NO: 228, SEQ ID NO: 236, SEQ ID NO: 241, SEQ ID NO: 249, SEQ ID NO: 277, SEQ ID NO: 282, SEQ ID NO: 300, SEQ ID NO: 349, SEQ ID NO: 362, SEQ ID NO: 365, SEQ ID NO: 383, SEQ ID NO: 385, SEQ ID NO: 388, a 15 biological equivalent thereof, or a fragment thereof.
- In certain other preferred embodiments, a polypeptide having a LPXTG motif and covalently attached to the peptidoglycan layer, comprises an amino acid sequence chosen from one of SEQ ID NO: 228, SEQ ID NO: 236, SEQ ID NO: 249, SEQ, SEQ ID NO: 385, a biological equivalent thereof, or a fragment thereof; or a 20 polypeptide having a peptidoglycan binding motif and associated with the peptidoglycan layer comprises an amino acid sequence chosen from one of SEQ ID NO: 240, SEQ ID NO: 264, SEQ ID NO: 325, a biological equivalent thereof, or a fragment thereof.
- 25 In another preferred embodiment, a polypeptide having a signal sequence and a C-terminal Tyrosine or Phenylalanine amino acid comprises an amino acid sequence chosen from one of SEQ ID NO:226, SEQ ID NO:254, SEQ ID NO:289, SEQ ID NO:312, SEQ ID NO:321, SEQ ID NO: 340, SEQ ID NO:402, a biological equivalent thereof, or a fragment thereof.
- 30 In yet another preferred embodiment, a polypeptide having a tripeptide RGD sequence that potentially is involved in cell attachment comprises an amino acid sequence chosen from one of SEQ ID NO:216, SEQ ID NO:236, SEQ ID NO:281, SEQ ID NO:282, a biological equivalent thereof, or a fragment thereof.

In still another embodiment, a polypeptide identified by proteomics as surface exposed comprises an amino acid sequence chosen from one of SEQ ID NO: 229, SEQ ID NO: 231, SEQ ID NO: 232, SEQ ID NO: 261, SEQ ID NO: 279, SEQ ID NO: 281, SEQ ID NO: 282, SEQ ID NO: 284, SEQ ID NO: 286, SEQ ID NO: 289, SEQ ID NO: 306, SEQ ID NO: 318, SEQ ID NO: 331, SEQ ID NO: 343, SEQ ID NO: 346, SEQ ID NO: 351, SEQ ID NO: 366, SEQ ID NO: 371, SEQ ID NO: 374, SEQ ID NO: 377, SEQ ID NO: 379, SEQ ID NO: 387, SEQ ID NO: 391, SEQ ID NO: 393, SEQ ID NO: 394, SEQ ID NO: 395, SEQ ID NO: 397, SEQ ID NO: 420, a biological equivalent thereof, or a fragment thereof.

10 In yet another embodiment, a polypeptide identified by proteomics as membrane associated comprises an amino acid sequence chosen from one of SEQ ID NO: 592 through SEQ ID NO: 752, a biological equivalent thereof, or a fragment thereof.

In another aspect of the invention, the polypeptides are expressed and purified in a recombinant expression system. Thus, in certain embodiments, the invention provides a recombinant expression vector comprising a nucleotide sequence having at least about 95% identity to a nucleotide sequence chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO: 431 through SEQ ID NO: 591, a degenerate variant thereof, or a fragment thereof. In certain other  
20 embodiments, the polynucleotide is selected from the group consisting of DNA, chromosomal DNA, cDNA, RNA and antisense RNA. In another embodiment, the polynucleotide comprised within the vector further comprises heterologous nucleotide sequences. In other embodiments, the polynucleotide is operatively linked to one or more gene expression regulatory elements. In yet other embodiments, the polynucleotide encodes a polypeptide comprising an amino acid sequence having at  
25 least about 95% identity to an amino acid sequence chosen from one of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID NO: 592 through SEQ ID NO: 752, a biological equivalent thereof, or a fragment thereof. In a preferred embodiment, the vector is a plasmid.

30 In another aspect of the invention, there is provided a genetically engineered host cell, transfected, transformed or infected with a recombinant expression vector comprising a nucleotide sequence having at least about 95% identity to a nucleotide sequence chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID

NO: 431 through SEQ ID NO: 591, a degenerate variant thereof, or a fragment thereof. In a preferred embodiment, the host cell is a bacterial cell. In a further embodiment, the polynucleotide is expressed under suitable conditions to produce the encoded polypeptide, a biological equivalent thereof, or a fragment thereof, which is then recovered.

In other embodiments, the present invention provides an antibody specific for a *Streptococcus pneumoniae* polynucleotide chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO: 431 through SEQ ID NO: 591, a fragment thereof, a degenerate variant thereof, or an antibody specific for a *Streptococcus pneumoniae* polypeptide chosen from one of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID NO: 592 through SEQ ID NO: 752, a biological equivalent thereof, or a fragment thereof. In certain embodiments, the antibody is selected from the group consisting of monoclonal, polyclonal, chimeric, humanized and single chain. In a preferred embodiment, the antibody is monoclonal. In another preferred embodiment, the antibody is humanized.

The present invention further provides pharmaceutical compositions, in particular immunogenic compositions, for the prevention and/or treatment of bacterial infection. Thus, in one embodiment an immunogenic composition is provided comprising a polypeptide having an amino acid sequence chosen from one or more of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID NO: 592 through SEQ ID NO: 752, a biological equivalent thereof, or a fragment thereof. In certain embodiments, the composition further comprises a pharmaceutically acceptable carrier. In yet other embodiments, the immunogenic composition further comprises one or more adjuvants. In a preferred embodiment, the polypeptide of the immunogenic composition is further defined as a *Streptococcus pneumoniae* polypeptide having 0, 1 or 2 transmembrane domains, a *Streptococcus pneumoniae* polypeptide having 3 or more transmembrane domains, a *Streptococcus pneumoniae* polypeptide having an outer membrane domain or a periplasmic domain, a *Streptococcus pneumoniae* polypeptide having an inner membrane domain, a *Streptococcus pneumoniae* polypeptide identified by Blastp analysis, a *Streptococcus pneumoniae* polypeptide identified by Pfam analysis, a *Streptococcus pneumoniae* lipoprotein, a *Streptococcus pneumoniae* polypeptide having a LPXTG motif, wherein the polypeptide is covalently attached to the peptidoglycan layer, a

- Streptococcus pneumoniae* polypeptide having a peptidoglycan binding motif, wherein the polypeptide is associated with the peptidoglycan layer, a *Streptococcus pneumoniae* polypeptide having a signal sequence and a C-terminal Tyrosine or Phenylalanine amino acid, a *Streptococcus pneumoniae* polypeptide having a
- 5 tripeptide RGD sequence, a *Streptococcus pneumoniae* polypeptide identified by proteomics as surface exposed or a *Streptococcus pneumoniae* polypeptide identified by proteomics as membrane associated. In certain other embodiments, the immunogenic composition further comprises heterologous amino acids. In particular
- 10 embodiments, the polypeptide is a fusion polypeptide.
- In further embodiments, provided is an immunogenic composition comprising a polynucleotide having a nucleotide sequence chosen from one or more of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO: 431 through SEQ IDNO: 591, a degenerate variant thereof, or a fragment thereof and is comprised in an expression vector. In preferred embodiments, the vector is plasmid DNA. In another
- 15 embodiment, the polynucleotide comprises heterologous nucleotides. In still other embodiments, the polynucleotide is operatively linked to one or more gene expression regulatory elements. In yet other embodiments, the polynucleotide directs the expression of a neutralizing epitope of *Streptococcus pneumoniae*. In preferred embodiments, the immunogenic composition further comprises one or
- 20 more adjuvants.
- Also provided is a pharmaceutical composition comprising a polypeptide and a pharmaceutically acceptable carrier, wherein the polypeptide comprises an amino acid chosen from one of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID NO: 592 through SEQ ID NO: 752, a biological equivalent thereof, or a fragment thereof.
- 25 In preferred embodiments, the polypeptide is further defined as a *Streptococcus pneumoniae* polypeptide having 0, 1 or 2 transmembrane domains, a *Streptococcus pneumoniae* polypeptide having 3 or more transmembrane domains, a *Streptococcus pneumoniae* polypeptide having an outer membrane domain or a periplasmic domain, a *Streptococcus pneumoniae* polypeptide having an inner membrane
- 30 domain, a *Streptococcus pneumoniae* polypeptide identified by Blastp analysis, a *Streptococcus pneumoniae* polypeptide identified by Pfam analysis, a *Streptococcus pneumoniae* lipoprotein, a *Streptococcus pneumoniae* polypeptide having a LPXTG motif, wherein the polypeptide is covalently attached to the peptidoglycan layer, a

*Streptococcus pneumoniae* polypeptide having a peptidoglycan binding motif, wherein the polypeptide is associated with the peptidoglycan layer, a *Streptococcus pneumoniae* polypeptide having a signal sequence and a C-terminal Tyrosine or Phenylalanine amino acid, a *Streptococcus pneumoniae* polypeptide having a tripeptide RGD sequence, a *Streptococcus pneumoniae* polypeptide identified by proteomics as surface exposed or a *Streptococcus pneumoniae* polypeptide identified by proteomics as membrane associated. In certain embodiments, the polypeptide further comprises heterologous amino acids. In still other embodiments, the polypeptide is a fusion polypeptide.

10 In another embodiment, a method of immunizing against *Streptococcus pneumoniae* is provided comprising administering to a host an immunizing amount of an immunogenic composition comprising one or more polypeptides and a pharmaceutically acceptable carrier, wherein the polypeptide comprises an amino acid sequence chosen from one or more of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID NO: 592 through SEQ ID NO: 752, a biological equivalent thereof, or a fragment thereof. In certain embodiments, the polypeptide is a fusion polypeptide. In other embodiments, the method further comprises administering an adjuvant.

Other embodiments of the invention provide a DNA chip comprising an array of polynucleotides, wherein at least one of the polynucleotides comprise a nucleotide sequence chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO: 431 through SEQ ID NO: 591, a complement thereof, a degenerate variant thereof, or a fragment thereof.

Also provided is a protein chip comprising an array of polypeptides, wherein at least one of the polypeptides comprises an amino acid sequence chosen from one of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID NO: 592 through SEQ ID NO: 752, a biological equivalent thereof, or a fragment thereof.

The invention further provides methods of detecting *Streptococcus pneumoniae* polynucleotides and polypeptides as well as kits for diagnosing *Streptococcus pneumoniae* infection.

30 Other embodiments provide a method for the detection and/or identification of *Streptococcus pneumoniae* in a biological sample comprising contacting the sample with an oligonucleotide probe of a polynucleotide comprising the nucleotide sequence chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID

NO: 431 through SEQ ID NO: 591, a degenerate variant thereof, or a fragment thereof, under conditions permitting hybridization and detecting the presence of hybridization complexes in the sample, wherein hybridization complexes indicate the presence of *Streptococcus pneumoniae* in the sample.

- 5 Still other embodiments provide a method for the detection and/or identification of *Streptococcus pneumoniae* in a biological sample comprising a nucleotide sequence chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO: 431 through SEQ ID NO: 591, a degenerate variant thereof, or a fragment thereof, in the presence of nucleotides and a polymerase enzyme under  
10 conditions permitting primer extension and detecting the presence of primer extension products in the sample, wherein extension products indicate the presence of *Streptococcus pneumoniae* in the sample.

- Further embodiments provide a method for the detection and/or identification of *Streptococcus pneumoniae* in a biological sample comprising contacting the  
15 sample with an antibody specific for a polypeptide comprising an amino acid sequence chosen from one of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID NO: 592 through SEQ ID NO: 752, a biological equivalent thereof, or a fragment thereof, under conditions permitting immune complex formation and detecting the presence of immune complexes in the sample, wherein immune complexes indicate  
20 the presence of *Streptococcus pneumoniae* in the sample.

- In certain embodiments, provided is a method for the detection and/or identification of antibodies to *Streptococcus pneumoniae* in a biological sample comprising contacting the sample with a polypeptide comprising an amino acid sequence chosen from one of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID  
25 NO: 592 through SEQ ID NO: 752, a biological equivalent thereof, or a fragment thereof, under conditions permitting immune complex formation and detecting the presence of immune complexes in the sample, wherein immune complexes indicate the presence of *Streptococcus pneumoniae* in the sample.

- Other embodiments of the invention provide a kit comprising a container  
30 containing an isolated polynucleotide comprising a nucleotide sequence chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO: 431 through SEQ ID NO: 591, a degenerate variant thereof, or a fragment thereof. In a preferred embodiment, the polynucleotide is a primer or a probe, wherein when the

polynucleotide is a primer, the kit further comprises a container containing a polymerase. In another embodiment, the kit further comprises a container containing dNTP.

5 Provided further is a kit comprising a container containing an antibody that immunospecifically binds to a polypeptide comprising the amino acid sequence chosen from one of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID NO: 592 through SEQ ID NO: 752, a biological equivalent thereof, or a fragment thereof.

10 Provided also is a kit comprising a container containing an antibody that immunospecifically binds to a fusion polypeptide comprising at least the amino acid sequence chosen from one of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID NO: 592 through SEQ ID NO: 752, a biological equivalent thereof, or a fragment thereof.

15 In a preferred embodiment of the invention, provided is a genetically engineered host cell, transfected, transformed or infected with a recombinant expression vector comprising a nucleotide sequence having at least about 95% identity to a nucleotide sequence chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO: 431 through SEQ ID NO: 591, a degenerate variant thereof, or a fragment thereof under conditions suitable to produce one of the polypeptides of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID NO: 592 through SEQ ID NO: 752; and recovering the polypeptide.

20 Other features and advantages of the invention will be apparent from the following detailed description, from the preferred embodiments thereof, and from the claims.

25

#### DETAILED DESCRIPTION OF THE INVENTION

The invention described hereinafter addresses the need for *Streptococcus pneumoniae* immunogenic compositions that effectively prevent or treat most or all of the disease caused by serotypes of *Streptococcus pneumoniae*. The invention  
30 further addresses the need for methods of diagnosing *Streptococcus pneumoniae* infection. The present invention has identified novel *Streptococcus pneumoniae* open reading frames, hereinafter ORFs, which encode antigenic polypeptides. More particularly, the newly identified ORFs encode polypeptides that are secreted, exposed, membrane associated or surface localized on *Streptococcus pneumoniae*,



and thus serve as potential antigenic polypeptides in immunogenic compositions. Thus, in certain embodiments, the invention comprises *Streptococcus pneumoniae* polynucleotide ORFs encoding surface localized, exposed, secreted or membrane associated polypeptide antigens. The present invention therefore comprises in other  
5       embodiments, these polypeptides, preferably antigenic polypeptides, encoded by the *Streptococcus pneumoniae* ORFs.

In other embodiments, the invention comprises vectors comprising ORF sequences and host cells or animals transformed, transfected or infected with these vectors. The invention also comprises transcriptional gene products of  
10       *Streptococcus pneumoniae* ORFs, such as, for example, mRNA, antisense RNA, antisense oligonucleotides and ribozyme molecules, which can be used to inhibit or control growth of the microorganism. The invention relates also to methods of detecting these nucleic acids or polypeptides and kits for diagnosing *Streptococcus pneumoniae* infection. The invention also relates to pharmaceutical compositions, in  
15       particular immunogenic compositions, for the prevention and/or treatment of bacterial infection, in particular infection caused by or exacerbated by *Streptococcus pneumoniae*. In particular embodiments, the immunogenic compositions are used for the treatment or prevention of systemic diseases which are induced or exacerbated by *Streptococcus pneumoniae*. In other embodiments, the immunogenic  
20       compositions are used for the treatment or prevention of non-systemic diseases, particularly of the otitis media, which are induced or exacerbated by *Streptococcus pneumoniae*.

#### **A. IDENTIFYING ORFs WITHIN THE GENOMIC SEQUENCE OF *STREPTOCOCCUS PNEUMONIAE***

  
25

The invention is directed in particular embodiments to the identification of polynucleotides, more particularly ORFs, that encode *Streptococcus pneumoniae* polypeptides. The availability of complete bacterial genome sequences has begun to play an important role in the identification of candidate antigens through genomics,  
30       transcriptional profiling, and proteomics, coupled with the information processing capabilities of bioinformatics (McAttee *et al.*, 1998a; McAttee *et al.*, 1998b; Pizzia *et al.*, 2000; Sonnenberg and Belisle, 1997; Welding *et al.*, 1998; McAttee *et al.*, 1998c). Currently, no more than approximately 60% of all ORFs within a bacterial genome

have some match with a polypeptide whose function has been determined. This leaves approximately 40% of genomic ORFs uncharacterized. Thus, the inventors have analyzed the *Streptococcus pneumoniae* genome and utilized bioinformatic tools to identify novel ORFs encoding polypeptides of the present invention. In addition to genomic analysis, the inventors analyzed the *Streptococcus pneumoniae* membrane proteome component to identify novel and/or confirm ORFs encoding polypeptides of the present invention. As described below, the ORFs were analyzed for a variety of characteristics.

Specifically, an extensive genomic analysis was performed *in silico* of the *Streptococcus pneumoniae* type 4 genome from The Institute for Genomic Research (TIGR) using algorithms designed to identify genes that encode novel surface localized polypeptides or polypeptides with putative similarity to polypeptides of known interest in other organisms. Thus, a combined analysis of the *Streptococcus pneumoniae* genome, using a unique set of two ORF finder algorithms (i.e., GLIMMER, Salzberg *et al.*, 1998 and inventors' assignee's own program), produced 3,799 ORFs. The most stringent of the ORF finders; Glimmer, produced 2,022 ORFs, while the assignee's ORF finder produced the most with 3,798 ORFs. There were 2,021 ORFs identified by the two algorithms. The difference in results between the different ORF finders is primarily due to the particular start codons used by each program; however, Glimmer also incorporates some evaluation for a Shine-Dalgarno box and an interpolated Markov model. For the purposes here, all ORFs with common stop codons are given the same ORF designation and will be treated as if they are the same ORF. As used hereinafter, an ORF is defined as having one of three potential start site codons, ATG, GTG or TTG and one of three potential stop codons, TAA, TAG or TGA. The lower limit of amino acid length selected as a cutoff (e.g., ~74 amino acids) may also cause the algorithms to overlook some reading frames. However, these "true" reading frames become an increasingly rare event as the ORFs become shorter.

The initial annotation of the *Streptococcus pneumoniae* ORFs was performed using the Basic Local Alignment Search Tool (BLAST; version 2.0) Gapped search algorithm, Blastp, to identify homologous sequences (Altschul *et al.*, 1997). A cutoff 'e' value of anything  $< e^{-10}$  was considered significant. The non-redundant protein sequence database used for the homology searches consisted of GenBank, SWISS-

PROT (Bairoch and Apweiler, 2000), PIR (Barker *et al.*, 2001), and TREMBL (Bairoch and Apweiler, 2000); whose database sequences are updated daily. In the present invention, ORFs with a Blastp result of  $> e^{-10}$  are considered to be unique to *Streptococcus pneumoniae*. Alternate quantitative expression values other than Blastp 'e', e.g., percent identity, may also be used to compare database sequences with the *Streptococcus pneumoniae* ORFs of the present invention.

A keyword search of the entire BLAST results was carried out using known or suspected target genes for immunogenic compositions as well as words that identified the location of a protein or function.

Several parameters were used to determine grouping of the predicted *Streptococcus pneumoniae* polypeptides of the invention. For example, polypeptides destined for translocation across the cytoplasmic membrane encode a leader signal (also called signal sequence) composed of a central hydrophobic region flanked at the N-terminus by positively charged residues (Pugsley, 1993). A software program, called SignalP, which identifies signal peptides and their cleavage sites based on neural networks (Nielsen *et al.*, 1997), was used in the present invention to analyze the amino acid sequence of an ORF for such a signal peptide. The first 60 N-terminal amino acids of each ORF were analyzed by SignalP using the Gram-positive software database. The output generated four separate values, maximum C, maximum Y, maximum S, and mean S. The S-score, or signal region, is the probability of the position belonging to the signal peptide. The C-score, or cleavage site, is the probability of the position being the first in the mature protein. The Y-score is the geometric average of the C-score and a smoothed derivative of the S-score. A conclusion of either a Yes or No is given next to each score. If all four conclusions are Yes, then a 'YES' is listed for that ORF; if three of the conclusions are Yes, then a 'yes' is listed for that ORF; if two of the conclusions are Yes, then a 'maybe' is listed for that ORF; for all other cases, a 'no' is listed for that ORF.

To predict polypeptide localization in bacteria, the software program PSORT was used (Nakai, 1991). PSORT predicts localization of polypeptides to the 'cytoplasm', 'periplasm', and/or 'cytoplasmic membrane' for Gram-positive bacteria, as well as 'outer membrane' for Gram-negative bacteria. Transmembrane (TM) domains of polypeptides were analyzed using the software program TopPred II (Cserzo *et al.*, 1997).

The Hidden Markov Model (HMM) Pfam database (Bateman, 2000) was used to identify *Streptococcus pneumoniae* proteins that may belong to an existing protein family. Keyword searching of this output was further used to help identify additional candidate antigens that may have been missed by the BLAST search criteria.

5        A computer algorithm, called HMM Lipo, was developed by inventors' assignee to predict lipoproteins using approximately 131 biologically proven bacterial lipoproteins. The protein sequence from the start of the protein to the cysteine amino acid, plus the next two additional amino acids, was used to generate the HMM (Eddy and Markov, 1996)

10        The inventor's assignee's also developed a HMM using approximately 70 known prokaryotic proteins containing the LPXTG cell wall sorting signal, to predict cell wall proteins that are anchored to the peptidoglycan layer (Mazmanian *et al.*, 1999; Navarre and Schneewind, 1999). The model used not only the LPXTG sequence, but also included two features of the downstream sequence, first the  
15        hydrophobic transmembrane domain and secondly, the positively charged carboxy terminus. There are also a number of proteins that interact, non-covalently, with the peptidoglycan layer and are distinct from the LPXTG protein class described above. These proteins seem to have a consensus sequence at their carboxy terminus (Koebnik, 1995). The inventors therefore developed and used a HMM of this region  
20        to identify any *Streptococcus pneumoniae* that may fall into this class of proteins.

*Streptococcus pneumoniae* ORFs encoding surface localized, exposed, or membrane associated polypeptides were also identified by proteomics (see, Example 3). This proteomic analysis confirmed many of the *Streptococcus pneumoniae* ORFs identified by the above genomic analysis and further identified novel *Streptococcus*  
25        *pneumoniae* ORFs encoding membrane associated polypeptides.

      The following Tables (*i.e.*, Tables 1-12) represent 12 groups into which the ORFs identified according to the above characteristics of present invention have been classified. Thus, all of the groups described below are ORFs comprised within the *Streptococcus pneumoniae* genome and identified as encoding putative surface  
30        localized, exposed, membrane associated or secreted polypeptides. These groups are not meant to limit the scope of the present invention, as analysis of additional ORF characteristics also are contemplated. These additional characteristics, *e.g.*, RGD sequence, may serve to further expand the total number of ORF groupings or

to parse the presently identified ORFs into more defined groups, broader groups, narrower groups or group subsets. In addition, some ORFs will meet the criteria of more than one category, and will therefore appear in more than one of the following groups.

- 5        Listed in Table 1 are ORFs that comprise a cytoplasmic membrane signal sequence (*i.e.*, a SignalP value of 'YES') and have one or fewer membrane spanning domains (MSD), as defined by the TopPred II program. Thirteen ORFs are found that match these criteria and are considered to be surface exposed.

**Table 1.** ORFs encoding surface exposed polypeptides, SignalP value = 'YES' and  $\leq 1$  MSDs.

SEQ ID	ORF
11	190
17	403
23	469
39	790
50	935
70	1143
83	1475
91	1568
97	1724
128	2271
148	2621
179	3212
209	3600

5

Listed in Table 2 are ORFs that comprise a cytoplasmic membrane signal sequence (*i.e.*, a SignalP value of 'YES') and an outer membrane (OM) or periplasmic (Peri) prediction value when analyzed via the program Psort. Five ORFs are found that match these criteria and are considered to be surface exposed.

10

**Table 2.** ORFs encoding surface exposed polypeptides, a SignalP value = 'YES' and a Psort value of 'OM or Peri'.

SEQ ID	ORF
23	469
39	790
50	935
125	2228
179	3212

Listed in Table 3 are ORFs that comprise a cytoplasmic membrane signal sequence (*i.e.*, a SignalP value of 'YES') and have 2 or more membrane spanning domains (MSD), as defined by the TopPred II program. Twenty two ORFs are found that match these criteria and are considered to be surface exposed.

5

**Table 3.** ORFs encoding surface exposed polypeptides, a SignalP = 'YES' and  $\leq 1$  MSDs.

SEQ ID	ORF
11	190
13	339
17	403
23	469
34	640
39	790
50	935
70	1143
73	1207
83	1475
91	1568
97	1724
106	1947
121	2196
125	2228
126	2234
128	2271
148	2621
179	3212
187	3361
192	3384
209	3600

10

Listed in Table 4 are ORFs that comprise at least 3 of 4 SignalP values (*i.e.*, a SignalP value of 'yes') and have 2 or more membrane spanning domains (MSD), as defined by the TopPred II program. Forty-nine ORFs are found that match these criteria and are considered to be surface exposed.

**Table 4.** ORFs encoding surface exposed polypeptides, a SignalP = 'yes' and  $\geq 2$  MSDs.

SEQ ID	ORF	SEQ ID	ORF
2	72	129	2304
6	94	133	2350
10	141	140	2470
14	356	145	2594
22	462	146	2613
28	597	152	2676
29	598	156	2838
36	715	168	3072
37	716	175	3141
40	823	180	3256
46	885	184	3340
47	904	188	3369
48	916	190	3373
56	989	194	3386
59	998	203	3558
71	1178	211	3631
77	1339	213	3770
80	1412	215	3799
81	1437		
86	1493		
87	1528		
88	1530		
93	1623		
99	1816		
101	1849		
102	1863		
105	1904		
112	2026		
114	2061		
115	2112		
120	2195		

5

Keyword search of the Blastp data for putative surface exposed proteins produced 119 ORFs and are listed in Table 5.



**Table 5.** ORFs encoding surface exposed polypeptides identified by keyword search of Blastp data.

SEQ ID	ORF	SEQ ID	ORF	SEQ ID	ORF	SEQ ID	ORF
1	51	88	1530	158	2847	213	3770
2	72	90	1560	159	2894	214	3789
7	113	94	1630	160	2969		
10	141	95	1632	161	2975		
12	304	96	1710	162	2979		
16	378	98	1765	163	2980		
20	410	100	1835	165	3039		
24	493	103	1864	166	3040		
27	580	105	1904	167	3060		
30	607	107	1966	169	3079		
31	612	108	1999	172	3107		
32	624	109	2001	173	3115		
33	639	112	2026	176	3167		
34	640	113	2027	177	3198		
35	703	115	2112	178	3209		
38	772	117	2132	180	3256		
40	823	118	2191	181	3262		
42	838	122	2198	182	3298		
43	854	123	2201	184	3340		
44	855	124	2215	185	3346		
48	916	127	2239	186	3349		
51	945	129	2304	188	3369		
53	979	131	2329	189	3372		
59	998	132	2348	191	3378		
60	1013	133	2350	193	3385		
61	1048	134	2352	196	3457		
65	1072	135	2354	197	3473		
67	1104	136	2385	198	3479		
68	1117	138	2431	199	3480		
69	1141	139	2452	200	3487		
70	1143	141	2488	201	3493		
71	1178	144	2591	202	3494		
75	1244	146	2613	204	3568		
76	1267	147	2615	205	3576		
77	1339	151	2661	206	3578		
78	1350	152	2676	207	3584		
79	1410	154	2734	208	3585		
80	1412	155	2814	210	3627		
87	1528	157	2845	212	3669		

HMM Pfam analysis helps identify ORFs encoding proteins with domains or amino acid patterns similar to proteins that belong to an existing protein family. Keyword search of the Pfam family classification for potential surface exposed proteins produced 52 ORFs and are listed in Table 6.

5

**Table 6.** ORFs encoding surface exposed polypeptides identified by HMM Pfam analysis.

SEQ ID	ORF	SEQ ID	ORF
4	79	160	2969
18	404	162	2979
19	406	163	2980
41	828	164	2983
45	869	165	3039
55	983	166	3040
57	992	169	3079
58	996	171	3083
63	1064	174	3140
64	1070	176	3167
66	1097	180	3256
72	1179	182	3298
74	1220	183	3327
89	1559	184	3340
92	1572	186	3349
104	1868	188	3369
111	2025	189	3372
116	2129	195	3413
119	2193	198	3479
128	2271	199	3480
137	2400	205	3576
142	2499	212	3669
143	2543	213	3770
149	2642		
151	2661		
152	2676		
153	2678		
157	2845		
159	2894		

An algorithm called HMM Lipo was developed for use in the present invention. The HMM Lipo program predicts lipoproteins using approximately 131 biologically proven bacterial lipoproteins. HMM Lipo identified 16 ORFs that are putative lipoproteins and are listed in Table 7.

5

**Table 7.** ORFs encoding surface exposed lipoproteins.

SEQ ID	ORF
3	75
8	132
9	140
13	339
21	423
26	502
34	640
62	1059
67	1104
85	1479
134	2352
147	2615
150	2655
168	3072
170	3081
173	3115

10 The inventors developed an HMM using approximately 70 known prokaryotic polypeptides containing the LPXTG cell wall sorting signal. Thus, this HMM was used to predict cell wall polypeptides that are anchored to the peptidoglycan layer. Listed in Table 8 are 4 ORFs predicted to have the LPXTG motif and are classified as proteins that might be targeted by sortase.

**Table 8.** ORFs encoding surface exposed polypeptides anchored to the peptidoglycan layer.

SEQ ID	ORF
13	339
21	423
34	640
170	3081

5

In addition, listed in Table 9 are 3 ORFs predicted by HMM PGB analysis to encode polypeptides potentially binding to the peptidoglycan layer in a manner independently of the sortase.

10

**Table 9.** ORFs encoding surface exposed polypeptides non-covalently anchored to the peptidoglycan layer.

SEQ ID	ORF
25	494
49	927
110	2012

15

ORFs that give a SignalP value of 'YES' and whose carboxy terminal amino acid is either a Phenylalanine or Tyrosine are considered to be surface exposed.

20 Listed in Table 10 are 7 ORFs matching these criteria.

**Table 10.** ORFs encoding surface exposed polypeptides, a cytoplasmic membrane signal sequence (*i.e.*, SignalP = 'YES') and a C-terminal Phe or Tyr amino acid.

5

SEQ ID	ORF
11	190
39	790
73	1207
97	1724
106	1947
125	2228
187	3361

10

Twenty eight *Streptococcus pneumoniae* ORFs were additionally identified by proteomics as encoding membrane associated polypeptides and are listed in Table 11. The ORFs listed in Table 11 further support the *Streptococcus pneumoniae* ORFs identified by the genomic mining algorithms described above (*i.e.*, ORFs encoding surface localized, secreted, or exposed polypeptides; Tables 1-10).

15

**Table 11.** *Streptococcus pneumoniae* ORFs confirmed by proteomics as surface exposed.

SEQ ID	ORF
14	356
16	378
17	403
46	885
64	1070
66	1097
67	1104
69	1141
71	1178
74	1220
91	1568
103	1864
116	2129
128	2271
131	2329
136	2385
151	2661
156	2838
159	2894
162	2979
164	2983
172	3107
176	3167
178	3209
179	3212
180	3256
182	3298
205	3576

5

Finally, 161 novel *Streptococcus pneumoniae* ORFs were identified by proteomics as encoding membrane associated polypeptides and are listed in Table 12.

**Table 12.** *Streptococcus pneumoniae* ORFs identified by proteomics as membrane associated.

SEQ ID	ORF	SEQ ID	ORF	SEQ ID	ORF	SEQ ID	ORF
431	64	463	357	495	1344	527	2284
432	120	464	390	496	1347	528	2315
433	121	465	431	497	1356	529	2317
434	152	466	434	498	1417	530	2318
435	153	467	436	499	1465	531	2319
436	156	468	439	500	1477	532	2320
437	159	469	513	501	1515	533	2372
438	160	470	515	502	1527	534	2374
439	163	471	583	503	1565	535	2376
440	164	472	633	504	1601	536	2387
441	166	473	683	505	1606	537	2394
442	172	474	686	506	1641	538	2410
443	174	475	720	507	1770	539	2425
444	175	476	726	508	1773	540	2443
445	178	477	818	509	1774	541	2451
446	180	478	861	510	1785	542	2454
447	181	479	863	511	1803	543	2508
448	183	480	960	512	1817	544	2513
449	186	481	1004	513	1823	545	2542
450	188	482	1037	514	1847	546	2558
451	189	483	1049	515	1917	547	2568
452	192	484	1054	516	1923	548	2575
453	194	485	1061	517	1964	549	2587
454	199	486	1082	518	1970	550	2754
455	268	487	1105	519	2039	551	2800
456	269	488	1111	520	2041	552	2839
457	294	489	1175	521	2047	553	2892
458	296	490	1248	522	2058	554	2906
459	298	491	1262	523	2068	555	2958
460	301	492	1266	524	2130	556	2963
461	316	493	1312	525	2251	557	3021
462	320	494	1314	526	2282	558	3048

**Table 12.** *Streptococcus pneumoniae* ORFs identified by proteomics as membrane associated.

SEQ ID	ORF	SEQ ID	ORF	SEQ ID	ORF	SEQ ID	ORF
559	3065	569	3248	579	3552	589	3739
560	3095	570	3303	580	3555	590	3766
561	3111	571	3331	581	3560	591	3778
562	3125	572	3367	582	3564		
563	3151	573	3410	583	3566		
564	3153	574	3446	584	3632		
565	3161	575	3454	585	3653		
566	3178	576	3525	586	3714		
567	3180	577	3538	587	3732		
568	3234	578	3540	588	3735		

5

As further contemplated in the present invention, *Streptococcus pneumoniae* ORFs are searched and evaluated for other important characteristics. For example, proteins that contain the Arg-Gly-Asp (RGD) attachment motif, together with integrins that serve as their receptor, constitute a major recognition system for cell adhesion, and thus are putative *Streptococcus pneumoniae* polypeptide antigens. Four *Streptococcus pneumoniae* ORFs, i.e., ORF 51, ORF 423, ORF 1097 and ORF 1104, have been identified as having a tripeptide RGD sequence that potentially is involved in cell attachment.

ORFs RGD recognition is one mechanism used by microbes to gain entry into eukaryotic tissues (Stockbauer *et al.*, 1999; Isberg and Nhlieu, 1994). However, not all RGD-containing proteins mediate cell attachment. It has been shown that RGD-containing peptides with a proline at the carboxy end (RGDP) are inactive in cell attachment assays (Pierschbacher and Rouslahti, 1987) and are excluded. A tandem repeat finder (Benson, 1999) may also be used, as has been used to identify ORFs containing repeated DNA sequences such as those found in MSCRAMMs (Foster and Hook, 1998) and phase variable surface proteins of *Neisseria meningitidis* (Parkhill *et al.*, 2000).

The present inventors also have used the Geanfammer software to cluster proteins into homologous families (Park and Teichmann, 1998). Preliminary analysis



of the family classes has provided novel ORFs within a vaccine candidate cluster as well as defining potential protein function.

The ORFs listed in Table 13, were identified by analysis of the *Streptococcus pneumoniae* genome. A total of 215 ORFs were identified based on the analysis criteria described above and listed in Tables 1-10. The 215 ORFs identified are listed vertically in Table 13 (column 1). The nucleotide SEQ ID NOS: 1 through SEQ ID NOS: 215 (column 2) and the encoded polypeptide SEQ ID NOS: 216 through SEQ ID NOS: 430 (column 3) are listed horizontally to their respective ORF. For example, in Table 13, ORF 51 has the nucleotide sequence of SEQ ID NO:1 and the encoded polypeptide has the amino acid sequence of SEQ ID NO: 216, ORF 72 has nucleotide SEQ ID NO:2 and encoded polypeptide SEQ ID NO: 217, etc.

Proteomic analysis identified twenty eight ORFs (see, Table 11) already listed in Table 13 (e.g., SEQ ID NO: 14, SEQ ID NO:16, SEQ ID NO:27, etc.) Proteomic analysis further identified 161 novel ORFs encoding membrane associated proteins (see, Table 12). These 161 novel ORFs identified by proteomics as membrane associated are listed vertically in Table 14 (column 1). The nucleotide SEQ ID NOS: 431 through SEQ ID NO: 591 (column 2) and the encoded polypeptide SEQ ID NOS: 592 through 752 (column 3) are listed horizontally to their respective ORF.

Table 13. *Streptococcus Pneumoniae* open reading frames (ORFs)

ORF	Nucleotide SEQ ID NO	Polypeptide SEQ ID NO
51	1	216
72	2	217
75	3	218
79	4	219
86	5	220
94	6	221
113	7	222
132	8	223
140	9	224
141	10	225
190	11	226
304	12	227
339	13	228
356	14	229
370	15	230
378	16	231
403	17	232
404	18	233
406	19	234
410	20	235
423	21	236
462	22	237
469	23	238
493	24	239
494	25	240
502	26	241
580	27	242
597	28	243
598	29	244
607	30	245
612	31	246
624	32	247
639	33	248
640	34	249
703	35	250
715	36	251
716	37	252
772	38	253
790	39	254
823	40	255

Table 13. *Streptococcus Pneumoniae* open reading frames (ORFs)

ORF	Nucleotide SEQ ID NO	Polypeptide SEQ ID NO
823	40	255
828	41	256
838	42	257
854	43	258
855	44	259
869	45	260
885	46	261
904	47	262
916	48	263
927	49	264
935	50	265
945	51	266
965	52	267
979	53	268
980	54	269
983	55	270
989	56	271
992	57	272
996	58	273
998	59	274
1013	60	275
1048	61	276
1059	62	277
1064	63	278
1070	64	279
1072	65	280
1097	66	281
1104	67	282
1117	68	283
1141	69	284
1143	70	285
1178	71	286
1179	72	287
1207	73	288
1220	74	289
1244	75	290
1267	76	291
1339	77	292
1350	78	293
1410	79	294

Table 13. *Streptococcus Pneumoniae* open reading frames (ORFs)

ORF	Nucleotide SEQ ID NO	Polypeptide SEQ ID NO
1412	80	295
1437	81	296
1459	82	297
1475	83	298
1476	84	299
1479	85	300
1493	86	301
1528	87	302
1530	88	303
1559	89	304
1560	90	305
1568	91	306
1572	92	307
1623	93	308
1630	94	309
1632	95	310
1710	96	311
1724	97	312
1765	98	313
1816	99	314
1835	100	315
1849	101	316
1863	102	317
1864	103	318
1868	104	319
1904	105	320
1947	106	321
1966	107	322
1999	108	323
2001	109	324
2012	110	325
2025	111	326
2026	112	327
2027	113	328
2061	114	329
2112	115	330
2129	116	331
2132	117	332
2191	118	333
2193	119	334

Table 13. *Streptococcus Pneumoniae* open reading frames (ORFs)

ORF	Nucleotide SEQ ID NO	Polypeptide SEQ ID NO
2195	120	335
2196	121	336
2198	122	337
2201	123	338
2215	124	339
2228	125	340
2234	126	341
2239	127	342
2271	128	343
2304	129	344
2322	130	345
2329	131	346
2348	132	347
2350	133	348
2352	134	349
2354	135	350
2385	136	351
2400	137	352
2431	138	353
2452	139	354
2470	140	355
2488	141	356
2499	142	357
2543	143	358
2591	144	359
2594	145	360
2613	146	361
2615	147	362
2621	148	363
2642	149	364
2655	150	365
2661	151	366
2676	152	367
2678	153	368
2734	154	369
2814	155	370
2838	156	371
2845	157	372
2847	158	373
2894	159	374

Table 13. *Streptococcus Pneumoniae* open reading frames (ORFs)

ORF	Nucleotide SEQ ID NO	Polypeptide SEQ ID NO
2969	160	375
2975	161	376
2979	162	377
2980	163	378
2983	164	379
3039	165	380
3040	166	381
3060	167	382
3072	168	383
3079	169	384
3081	170	385
3083	171	386
3107	172	387
3115	173	388
3140	174	389
3141	175	390
3167	176	391
3198	177	392
3209	178	393
3212	179	394
3256	180	395
3262	181	396
3298	182	397
3327	183	398
3340	184	399
3346	185	400
3349	186	401
3361	187	402
3369	188	403
3372	189	404
3373	190	405
3378	191	406
3384	192	407
3385	193	408
3386	194	409
3413	195	410
3457	196	411
3473	197	412
3479	198	413
3480	199	414

Table 13. *Streptococcus Pneumoniae* open reading frames (ORFs)

ORF	Nucleotide SEQ ID NO	Polypeptide SEQ ID NO
3487	200	415
3493	201	416
3494	202	417
3558	203	418
3568	204	419
3576	205	420
3578	206	421
3584	207	422
3585	208	423
3600	209	424
3627	210	425
3631	211	426
3669	212	427
3770	213	428
3789	214	429
3799	215	430

Table 14. *Streptococcus Pneumoniae* open reading frames (ORFs)

ORF	Nucleotide SEQ ID NO	Polypeptide SEQ ID NO
64	431	592
120	432	593
121	433	594
152	434	595
153	435	596
156	436	597
159	437	598
160	438	599
163	439	600
164	440	601
166	441	602
172	442	603
174	443	604
175	444	605
178	445	606
180	446	607
181	447	608
183	448	609
186	449	610
188	450	611
189	451	612
192	452	613
194	453	614
199	454	615
268	455	616
269	456	617
294	457	618
296	458	619
298	459	620
301	460	621
316	461	622
320	462	623
357	463	624
390	464	625
431	465	626
434	466	627
436	467	628



Table 14. *Streptococcus Pneumoniae* open reading frames (ORFs)

ORF	Nucleotide SEQ ID NO	Polypeptide SEQ ID NO
439	468	629
513	469	630
515	470	631
583	471	632
633	472	633
683	473	634
686	474	635
720	475	636
726	476	637
818	477	638
861	478	639
863	479	640
960	480	641
1004	481	642
1037	482	643
1049	483	644
1054	484	645
1061	485	646
1082	486	647
1105	487	648
1111	488	649
1175	489	650
1248	490	651
1262	491	652
1266	492	653
1312	493	654
1314	494	655
1344	495	656
1347	496	657
1356	497	658
1417	498	659
1465	499	660
1477	500	661
1515	501	662
1527	502	663
1565	503	664
1601	504	665

Table 14. *Streptococcus Pneumoniae* open reading frames (ORFs)

ORF	Nucleotide SEQ ID NO	Polypeptide SEQ ID NO
1606	505	666
1641	506	667
1770	507	668
1773	508	669
1774	509	670
1785	510	671
1803	511	672
1817	512	673
1823	513	674
1847	514	675
1917	515	676
1923	516	677
1964	517	678
1970	518	679
2039	519	680
2041	520	681
2047	521	682
2058	522	683
2068	523	684
2130	524	685
2251	525	686
2282	526	687
2284	527	688
2315	528	689
2317	529	690
2318	530	691
2319	531	692
2320	532	693
2372	533	694
2374	534	695
2376	535	696
2387	536	697
2394	537	698
2410	538	699
2425	539	700
2443	540	701
2451	541	702

Table 14. *Streptococcus Pneumoniae* open reading frames (ORFs)

ORF	Nucleotide SEQ ID NO	Polypeptide SEQ ID NO
2454	542	703
2508	543	704
2513	544	705
2542	545	706
2558	546	707
2568	547	708
2575	548	709
2587	549	710
2754	550	711
2800	551	712
2839	552	713
2892	553	714
2906	554	715
2958	555	716
2963	556	717
3021	557	718
3048	558	719
3065	559	720
3095	560	721
3111	561	722
3125	562	723
3151	563	724
3153	564	725
3161	565	726
3178	566	727
3180	567	728
3234	568	729
3248	569	730
3303	570	731
3331	571	732
3367	572	733
3410	573	734
3446	574	735
3454	575	736
3525	576	737
3538	577	738
3540	578	739

Table 14. *Streptococcus Pneumoniae* open reading frames (ORFs)

ORF	Nucleotide SEQ ID NO	Polypeptide SEQ ID NO
3552	579	740
3555	580	741
3560	581	742
3564	582	743
3566	583	744
3632	584	745
3653	585	746
3714	586	747
3732	587	748
3735	588	749
3739	589	750
3766	590	751
3778	591	752

**B. *STREPTOCOCCUS PNEUMONIAE* ORF POLYNUCLEOTIDES ENCODING SURFACE EXPOSED POLYPEPTIDES**

Isolated and purified *Streptococcus pneumoniae* ORF polynucleotides of the present invention are contemplated for use in the production of *Streptococcus pneumoniae* polypeptides. More specifically, in certain embodiments, the ORFs encode *Streptococcus pneumoniae* surface localized, exposed, membrane associated or secreted polypeptides, particularly antigenic polypeptides. Thus, in one aspect, the present invention provides isolated and purified polynucleotides (ORFs) that encode *Streptococcus pneumoniae* surface localized, exposed, membrane associated or secreted polypeptides. In particular embodiments, a polynucleotide of the present invention is a DNA molecule, wherein the DNA may be genomic DNA, chromosomal DNA, plasmid DNA or cDNA. In a preferred embodiment, a polynucleotide of the present invention is a recombinant polynucleotide, which encodes a *Streptococcus pneumoniae* polypeptide comprising an amino acid sequence that has at least 95% identity to an amino acid sequence of one of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID NO: 592 through SEQ ID NO: 752, or a fragment thereof. In another embodiment, an isolated and purified ORF polynucleotide comprises a nucleotide sequence that has at least 95% identity

to one of the ORF nucleotide sequences of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO: 431 through SEQ ID NO: 591, a degenerate variant thereof, or a complement thereof. In a preferred embodiment, an ORF polynucleotide of one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO: 431 through SEQ ID NO: 591 is comprised in a plasmid vector and expressed in a prokaryotic host cell.

As used hereinafter, the term "polynucleotide" means a sequence of nucleotides connected by phosphodiester linkages. Polynucleotides are presented hereinafter in the direction from the 5' to the 3' direction. A polynucleotide of the present invention can comprise from about 10 to about several hundred thousand base pairs. Preferably, a polynucleotide comprises from about 10 to about 3,000 base pairs. Preferred lengths of particular polynucleotide are set forth hereinafter.

A polynucleotide of the present invention can be a deoxyribonucleic acid (DNA) molecule, a ribonucleic acid (RNA) molecule, or analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA. Where a polynucleotide is a DNA molecule, that molecule can be a gene, a cDNA molecule or a genomic DNA molecule. Nucleotide bases are indicated hereinafter by a single letter code: adenine (A), guanine (G), thymine (T), cytosine (C), inosine (I) and uracil (U).

"Isolated" means altered "by the hand of man" from the natural state. If an "isolated" composition or substance occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living animal is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated," as the term is employed hereinafter.

Preferably, an "isolated" polynucleotide is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated *Streptococcus pneumoniae* nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. However, the

*Streptococcus pneumoniae* nucleic acid molecule can be fused to other protein encoding or regulatory sequences and still be considered isolated.

ORF polynucleotides of the present invention may be obtained, using standard cloning and screening techniques, from a cDNA library derived from mRNA.

5 Polynucleotides of the invention can also be obtained from natural sources such as genomic DNA libraries (e.g., a *Streptococcus pneumoniae* library) or can be synthesized using well known and commercially available techniques. Contemplated in the present invention, ORF polynucleotides will be obtained using *Streptococcus pneumoniae* type 3, type 14 or type 19F chromosomal DNA as the template.

10 The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences shown in SEQ ID NO:1 through SEQ ID NO:215 or SEQ ID NO: 431 through SEQ ID NO: 591 (and fragments thereof) due to degeneracy of the genetic code and thus encode the same *Streptococcus pneumoniae* polypeptide as that encoded by the nucleotide sequence shown SEQ ID NO:1 through SEQ ID  
15 NO:215 or SEQ ID NO: 431 through SEQ ID NO: 591.

Orthologues and allelic variants of the *Streptococcus pneumoniae* polynucleotides can readily be identified using methods well known in the art. Allelic variants and orthologues of the polynucleotides will comprise a nucleotide sequence that is typically at least about 70-75%, more typically at least about 80-85%, and  
20 most typically at least about 90-95% or more homologous to the nucleotide sequence shown in SEQ ID NO:1 through SEQ ID NO:215 or SEQ ID NO: 431 through SEQ ID NO: 591, or a fragment of these nucleotide sequences. Such nucleic acid molecules can readily be identified as being able to hybridize, preferably under stringent conditions, to the nucleotide sequence shown in SEQ ID NO:1 through SEQ ID  
25 NO:215 or SEQ ID NO: 431 through SEQ ID NO: 591, or a fragment of these nucleotide sequences.

Moreover, the polynucleotide of the invention can comprise only a fragment of the coding region of a *Streptococcus pneumoniae* polynucleotide or gene, such as a fragment of one of SEQ ID NO:1 through SEQ ID NO:215 or SEQ ID NO: 431  
30 through SEQ ID NO: 591. Preferably, such fragments are immunogenic fragments.

When the ORF polynucleotides of the invention are used for the recombinant production of *Streptococcus pneumoniae* polypeptides of the present invention, the polynucleotide may include the coding sequence for the mature polypeptide, by itself,

or the coding sequence for the mature polypeptide in reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro- protein sequence, or other fusion peptide portions. For example, a marker sequence which facilitates purification of the fused polypeptide can be linked to the coding sequence (see Gentz *et al.*, 1989, incorporated by reference hereinafter in its entirety). Thus, contemplated in the present invention is the preparation of polynucleotides encoding fusion polypeptides permitting His-tag purification of expression products. The polynucleotide may also contain non-coding 5' and 3' sequences, such as transcribed, non-translated sequences, splicing and polyadenylation signals.

Thus, a polynucleotide encoding a polypeptide of the present invention, including homologs and orthologs from species other than *Streptococcus pneumoniae*, may be obtained by a process which comprises the steps of screening an appropriate library under stringent hybridization conditions with a labeled probe having the sequence of one of SEQ ID NO:1 through SEQ ID NO:215 or SEQ ID NO:431 through SEQ ID NO: 591, a fragment thereof; and isolating full-length cDNA and genomic clones containing the polynucleotide sequence. Such hybridization techniques are well known to the skilled artisan. The skilled artisan will appreciate that, in many cases, an isolated cDNA sequence will be incomplete, in that the region coding for the polypeptide is cut short at the 5' end of the cDNA. This is a consequence of reverse transcriptase, an enzyme with inherently low "processivity" (a measure of the ability of the enzyme to remain attached to the template during the polymerization reaction), failing to complete a DNA copy of the mRNA template during 1st strand cDNA synthesis.

Thus, in certain embodiments, the polynucleotide sequence information provided by the present invention allows for the preparation of relatively short DNA (or RNA) oligonucleotide sequences having the ability to specifically hybridize to gene sequences of the selected polynucleotides disclosed hereinafter. The term "oligonucleotide" as used hereinafter is defined as a molecule comprised of two or more deoxyribonucleotides or ribonucleotides, usually more than three (3), and typically more than ten (10) and up to one hundred (100) or more (although preferably between twenty and thirty). The exact size will depend on many factors, which in turn depends on the ultimate function or use of the oligonucleotide. Thus, in

particular embodiments of the invention, nucleic acid probes of an appropriate length are prepared based on a consideration of a selected nucleotide sequence, e.g., a sequence such as that shown in SEQ ID NO:1 through SEQ ID NO:215 or SEQ ID NO: 431 through SEQ ID NO: 591. The ability of such nucleic acid probes to specifically hybridize to a polynucleotide encoding a *Streptococcus pneumoniae* polypeptide lends them particular utility in a variety of embodiments. Most importantly, the probes can be used in a variety of assays for detecting the presence of complementary sequences in a given sample.

In certain embodiments, it is advantageous to use oligonucleotide primers. These primers may be generated in any manner, including chemical synthesis, DNA replication, reverse transcription, or a combination thereof. The sequence of such primers is designed using a polynucleotide of the present invention for use in detecting, amplifying or mutating a defined segment of an ORF polynucleotide that encodes a *Streptococcus pneumoniae* polypeptide from prokaryotic cells using polymerase chain reaction (PCR) technology.

In certain embodiments, it is advantageous to employ a polynucleotide of the present invention in combination with an appropriate label for detecting hybrid formation. A wide variety of appropriate labels are known in the art, including radioactive, enzymatic or other ligands, such as avidin/biotin, which are capable of giving a detectable signal.

Polynucleotides which are identical or sufficiently identical to a nucleotide sequence contained in one of SEQ ID NO:1 through SEQ ID NO:215 or SEQ ID NO: 431 through SEQ ID NO: 591, or a fragment thereof, may be used as hybridization probes for cDNA and genomic DNA or as primers for a nucleic acid amplification (PCR) reaction, to isolate full-length cDNAs and genomic clones encoding polypeptides of the present invention and to isolate cDNA and genomic clones of other genes (including genes encoding homologs and orthologs from species other than *Streptococcus pneumoniae*) that have a high sequence similarity to the polynucleotide sequences set forth in of SEQ ID NO:1 through SEQ ID NO:215 or SEQ ID NO: 431 through SEQ ID NO: 591, or a fragment thereof. Typically these nucleotide sequences are from at least about 70% identical to at least about 95% identical to that of the reference polynucleotide sequence. The probes or primers will generally comprise at least 15 nucleotides, preferably, at least 30 nucleotides and



may have at least 50 nucleotides. Particularly preferred probes will have between 30 and 50 nucleotides.

There are several methods available and well known to those skilled in the art to obtain full-length cDNAs, or extend short cDNAs, for example those based on the method of Rapid Amplification of cDNA ends (RACE) (see, Frohman *et al.*, 1988). Recent modifications of the technique, exemplified by the Marathon™ technology (Clontech Laboratories Inc.) for example, have significantly simplified the search for longer cDNAs. In the Marathon™ technology, cDNAs have been prepared from mRNA extracted from a chosen tissue and an "adaptor" sequence ligated onto each end. Nucleic acid amplification (PCR) is then carried out to amplify the "missing" 5' end of the cDNA using a combination of gene specific and adaptor specific oligonucleotide primers. The PCR reaction is then repeated using "nested" primers, that is, primers designed to anneal within the amplified product (typically an adaptor specific primer that anneals further 3' in the adaptor sequence and a gene specific primer that anneals further 5' in the known gene sequence). The products of this reaction can then be analyzed by DNA sequencing and a full-length cDNA constructed either by joining the product directly to the existing cDNA to give a complete sequence, or carrying out a separate full-length PCR using the new sequence information for the design of the 5' primer.

To provide certain of the advantages in accordance with the present invention, a preferred nucleic acid sequence employed for hybridization studies or assays includes probe molecules that are complementary to at least a 10 to about 70 nucleotides long stretch of a polynucleotide that encodes a *Streptococcus pneumoniae* polypeptide, such as that shown in one of SEQ ID NO:216 through SEQ ID NO:430 or SEQ ID NO: 592 through SEQ ID NO: 752. A size of at least 10 nucleotides in length helps to ensure that the fragment will be of sufficient length to form a duplex molecule that is both stable and selective. Molecules having complementary sequences over stretches greater than 10 bases in length are generally preferred, though, in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of specific hybrid molecules obtained. One will generally prefer to design nucleic acid molecules having gene-complementary stretches of 25 to 40 nucleotides, 55 to 70 nucleotides, or even longer where desired. Such fragments can be readily prepared by, for example,

directly synthesizing the fragment by chemical means, by application of nucleic acid reproduction technology, such as the PCR technology of (U.S. Patent 4,683,202, incorporated hereinafter by reference) or by excising selected DNA fragments from recombinant plasmids containing appropriate inserts and suitable restriction enzyme sites.

In another aspect, the present invention contemplates an isolated and purified polynucleotide comprising a nucleotide sequence that is identical or complementary to a segment of at least 10 contiguous bases of one of SEQ ID NO:1 through SEQ ID NO:215 or SEQ ID NO: 431 through SEQ ID NO: 591, wherein the polynucleotide hybridizes to a polynucleotide that encodes a *Streptococcus pneumoniae* polypeptide. Preferably, the isolated and purified polynucleotide comprises a base sequence that is identical or complementary to a segment of at least 25 to about 70 contiguous bases of one of SEQ ID NO:1 through SEQ ID NO:215 or SEQ ID NO: 431 through SEQ ID NO: 591. For example, the polynucleotide of the invention can comprise a segment of bases identical or complementary to 40 or 55 contiguous bases of the disclosed nucleotide sequences.

Accordingly, a polynucleotide probe molecule of the invention can be used for its ability to selectively form duplex molecules with complementary stretches of the gene. Depending on the application envisioned, one will desire to employ varying conditions of hybridization to achieve varying degree of selectivity of the probe toward the target sequence (see Table 15 below). For applications requiring a high degree of selectivity, one will typically desire to employ relatively stringent conditions to form the hybrids. Of course, for some applications, for example, where one desires to prepare mutants employing a mutant primer strand hybridized to an underlying template or where one seeks to isolate a *Streptococcus pneumoniae* homologous polypeptide coding sequence from other cells, functional equivalents, or the like, less stringent hybridization conditions are typically needed to allow formation of the heteroduplex (see Table 15). Cross-hybridizing species can thereby be readily identified as positively hybridizing signals with respect to control hybridizations. Thus, hybridization conditions are readily manipulated, and thus will generally be a method of choice depending on the desired results.

Of course, for some applications, for example, where one desires to prepare mutants employing a mutant primer strand hybridized to an underlying template or

- where one seeks to isolate a homologous polypeptide coding sequence from other cells, functional equivalents, or the like, less stringent hybridization conditions are typically needed to allow formation of the heteroduplex. Cross-hybridizing species are thereby readily identified as positively hybridizing signals with respect to control
- 5 hybridizations. In any case, it is generally appreciated that conditions can be rendered more stringent by the addition of increasing amounts of formamide, which serves to destabilize the hybrid duplex in the same manner as increased temperature. Thus, hybridization conditions are readily manipulated, and thus will generally be a method of choice depending on the desired results.
- 10 The present invention also includes polynucleotides capable of hybridizing under reduced stringency conditions, more preferably stringent conditions, and most preferably highly stringent conditions, to polynucleotides described hereinafter. Examples of stringency conditions are shown in the table below: highly stringent conditions are those that are at least as stringent as, for example, conditions A-F;
- 15 stringent conditions are at least as stringent as, for example, conditions G-L; and reduced stringency conditions are at least as stringent as, for example, conditions M-R.

**Table 15**  
**Stringency Conditions**

Stringency Condition	Polynucleotide Hybrid	Hybrid Length (bp) <sup>1</sup>	Hybridization Temperature and Buffer <sup>1</sup>	Wash Temperature and Buffer <sup>1</sup>
A	DNA:DNA	> 50	65°C; 1xSSC -or- 42°C; 1xSSC, 50% formamide	65°C; 0.3xSSC
B	DNA:DNA	< 50	T <sub>B</sub> ; 1xSSC	T <sub>B</sub> ; 1xSSC
C	DNA:RNA	> 50	67°C; 1xSSC -or- 45°C; 1xSSC, 50% formamide	67°C; 0.3xSSC
D	DNA:RNA	< 50	T <sub>D</sub> ; 1xSSC	T <sub>D</sub> ; 1xSSC
E	RNA:RNA	> 50	70°C; 1xSSC -or- 50°C; 1xSSC, 50% formamide	70°C; 0.3xSSC
F	RNA:RNA	< 50	T <sub>F</sub> ; 1xSSC	T <sub>F</sub> ; 1xSSC
G	DNA:DNA	> 50	65°C; 4xSSC -or- 42°C; 4xSSC, 50% formamide	65°C; 1xSSC
H	DNA:DNA	< 50	T <sub>H</sub> ; 4xSSC	T <sub>H</sub> ; 4xSSC
I	DNA:RNA	> 50	67°C; 4xSSC -or- 45°C; 4xSSC, 50% formamide	67°C; 1xSSC
J	DNA:RNA	< 50	T <sub>J</sub> ; 4xSSC	T <sub>J</sub> ; 4xSSC
K	RNA:RNA	> 50	70°C; 4xSSC -or- 50°C; 4xSSC, 50% formamide	67°C; 1xSSC
L	RNA:RNA	< 50	T <sub>L</sub> ; 2xSSC	T <sub>L</sub> ; 2xSSC
M	DNA:DNA	> 50	50°C; 4xSSC -or- 40°C; 6xSSC, 50% formamide	50°C; 2xSSC
N	DNA:DNA	< 50	T <sub>N</sub> ; 6xSSC	T <sub>N</sub> ; 6xSSC

O	DNA:RNA	> 50	55°C; 4xSSC -or- 42°C; 6xSSC, 50% formamide	55°C; 2xSSC
P	DNA:RNA	< 50	T <sub>P</sub> ; 6xSSC	T <sub>P</sub> ; 6xSSC
Q	RNA:RNA	> 50	60°C; 4xSSC -or- 45°C; 6xSSC, 50% formamide	60°C; 2xSSC
R	RNA:RNA	< 50	T <sub>R</sub> ; 4xSSC	T <sub>R</sub> ; 4xSSC

(bp)<sup>1</sup>: The hybrid length is that anticipated for the hybridized region(s) of the hybridizing polynucleotides. When hybridizing a polynucleotide to a target polynucleotide of unknown sequence, the hybrid length is assumed to be that of the hybridizing polynucleotide. When polynucleotides of known sequence are hybridized, the hybrid length can be determined by aligning the sequences of the polynucleotides and identifying the region or regions of optimal sequence complementarity.

Buffer<sup>1</sup>: SSPE (1xSSPE is 0.15M NaCl, 10mM NaH<sub>2</sub>PO<sub>4</sub>, and 1.25mM EDTA, pH 7.4) can be substituted for SSC (1xSSC is 0.15M NaCl and 15mM sodium citrate) in the hybridization and wash buffers; washes are performed for 15 minutes after hybridization is complete.

T<sub>B</sub> through T<sub>R</sub>: The hybridization temperature for hybrids anticipated to be less than 50 base pairs in length should be 5-10°C less than the melting temperature (T<sub>m</sub>) of the hybrid, where T<sub>m</sub> is determined according to the following equations. For hybrids less than 18 base pairs in length, T<sub>m</sub>(°C) = 2(# of A + T bases) + 4(# of G + C bases). For hybrids between 18 and 49 base pairs in length, T<sub>m</sub>(°C) = 81.5 + 16.6(log<sub>10</sub>[Na<sup>+</sup>]) + 0.41(%G+C) - (600/N), where N is the number of bases in the hybrid, and [Na<sup>+</sup>] is the concentration of sodium ions in the hybridization buffer ([Na<sup>+</sup>] for 1xSSC = 0.165 M).

Additional examples of stringency conditions for polynucleotide hybridization are provided in Sambrook *et al.*, 1989, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, chapters 9 and 11, and Ausubel *et al.*, 1995, Current Protocols in Molecular Biology, eds., John Wiley & Sons, Inc., sections 2.10 and 6.3-6.4, incorporated hereinafter by reference.

In addition to the nucleic acid molecules encoding *Streptococcus pneumoniae* polypeptides described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid

encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire *Streptococcus pneumoniae* coding strand, or to only a fragment thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding a *Streptococcus pneumoniae* polypeptide.

The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues, e.g., the entire coding region of one of SEQ ID NO:1 through SEQ ID NO:215 or SEQ ID NO: 431 through SEQ ID NO: 591. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding a *Streptococcus pneumoniae* polypeptide. The term "noncoding region" refers to 5' and 3' sequences that flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

Given the coding strand sequence encoding the *Streptococcus pneumoniae* polypeptide disclosed hereinafter (e.g., one of SEQ ID NO:1 through SEQ ID NO:215 or SEQ ID NO: 431 through SEQ ID NO: 571), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of *Streptococcus pneumoniae* mRNA, but more preferably is an oligonucleotide which is antisense to only a fragment of the coding or noncoding region of *Streptococcus pneumoniae* mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of *Streptococcus pneumoniae* mRNA.

An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives

and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxoacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxoacetic acid methylester, uracil-5-oxoacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine.

Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a *Streptococcus pneumoniae* polypeptide to thereby inhibit expression of the polypeptide, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of an antisense nucleic acid molecule of the invention includes direct injection at a tissue site. Alternatively, an antisense nucleic acid molecule can be modified to target selected cells and then administered systemically. For example, for systemic administration, an antisense molecule can be modified such that it specifically binds to a receptor or an antigen expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecule to a peptide or an antibody which binds to a cell

surface receptor or antigen. The antisense nucleic acid molecule can also be delivered to cells using the vectors described hereinafter.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an  $\alpha$ -anomeric nucleic acid molecule. An  $\alpha$ -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual  $\gamma$ -units, the strands run parallel to each other (Gautier *et al.*, 1987). The antisense nucleic acid molecule can also comprise a 2'- $\alpha$ -methylribose nucleotide (Inoue *et al.*, 1987 (a)) or a chimeric RNA-DNA analogue (Inoue *et al.*, 1987(b)).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff and Gerlach, 1988)) can be used to catalytically cleave *Streptococcus pneumoniae* mRNA transcripts to thereby inhibit translation of *Streptococcus pneumoniae* mRNA. A ribozyme having specificity for a *Streptococcus pneumoniae*-encoding nucleic acid can be designed based upon the nucleotide sequence of a *Streptococcus pneumoniae* cDNA disclosed hereinafter (i.e., SEQ ID NO:1 through SEQ ID NO:215 or SEQ ID NO: 431 through SEQ ID NO: 591). For example, a derivative of a Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a *Streptococcus pneumoniae*-encoding mRNA. See, e.g., Cech *et al.* U.S. Patent 4,987,071 and Cech *et al.* U.S. Patent 5,116,742 both incorporated by reference. Alternatively, *Streptococcus pneumoniae* mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel and Szostak, 1993.

Alternatively *Streptococcus pneumoniae* gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the *Streptococcus pneumoniae* gene (e.g., the *Streptococcus pneumoniae* gene promoter and/or enhancers) to form triple helical structures that prevent transcription of the *Streptococcus pneumoniae* gene in target cells. See generally, Helene, 1991; Helene *et al.*, 1992; and Maher, 1992.



*Streptococcus pneumoniae* gene expression can also be inhibited using RNA interference (RNAi). This is a technique for post-transcriptional gene silencing (PTGS), in which target gene activity is specifically abolished with cognate double-stranded RNA (dsRNA). RNAi resembles in many aspects PTGS in plants and has been detected in many invertebrates including trypanosome, hydra, planaria, nematode and fruit fly (*Drosophila melanogaster*). It may be involved in the modulation of transposable element mobilization and antiviral state formation. RNAi in mammalian systems is disclosed in International Application WO 00/63364 which is incorporated by reference hereinafter in its entirety. Basically, dsRNA of at least about 600 nucleotides, homologous to the target is introduced into the cell and a sequence specific reduction in gene activity is observed.

### C. *STREPTOCOCCUS PNEUMONIAE* POLYPEPTIDES

In particular embodiments, the present invention provides isolated and purified *Streptococcus pneumoniae* polypeptides. Preferably, a *Streptococcus pneumoniae* polypeptide of the invention is a recombinant polypeptide. In certain embodiments, a *Streptococcus pneumoniae* polypeptide of the present invention comprises the amino acid sequence that has at least 95% identity to the amino acid sequence of one of SEQ ID NO:216 through SEQ ID NO:430 or SEQ ID NO: 592 through SEQ ID NO: 752 a biological equivalent thereof, or a fragment thereof.

A *Streptococcus pneumoniae* polypeptide according to the present invention encompasses a polypeptide that comprises: 1) the amino acid sequence shown in one of SEQ ID NO:216 through SEQ ID NO:430 or SEQ ID NO: 592 or SEQ ID NO: 752; 2) functional and non-functional naturally occurring variants or biological equivalents of *Streptococcus pneumoniae* polypeptides of SEQ ID NO:216 through SEQ ID NO:430 or SEQ ID NO: 592 through 752; 3) recombinantly produced variants or biological equivalents of *Streptococcus pneumoniae* polypeptides of SEQ ID NO:216 through SEQ ID NO:430 or SEQ ID NO: 592 through SEQ ID NO: 752; and 4) polypeptides isolated from organisms other than *Streptococcus pneumoniae* (orthologues of *Streptococcus pneumoniae* polypeptides.)

A biological equivalent or variant of a *Streptococcus pneumoniae* polypeptide according to the present invention encompasses 1) a polypeptide isolated from

*Streptococcus pneumoniae*; and 2) a polypeptide that contains substantially homology to a *Streptococcus pneumoniae* polypeptide.

- Biological equivalents or variants of *Streptococcus pneumoniae* include both functional and non-functional *Streptococcus pneumoniae* polypeptides. Functional biological equivalents or variants are naturally occurring amino acid sequence variants of a *Streptococcus pneumoniae* polypeptide that maintains the ability to elicit an immunological or antigenic response in a subject. Functional variants will typically contain only conservative substitution of one or more amino acids of one of SEQ ID NO:216 through SEQ ID NO:430 or SEQ ID NO: 592 through SEQ ID NO: 752, or substitution, deletion or insertion of non-critical residues in non-critical regions of the polypeptide (e.g., not in regions containing antigenic determinants or protective epitopes).

- The present invention further provides non-*Streptococcus pneumoniae* orthologues of *Streptococcus pneumoniae* polypeptides. Orthologues of *Streptococcus pneumoniae* polypeptides are polypeptides that are isolated from non-*Streptococcus pneumoniae* organisms and possess antigenic capabilities of the *Streptococcus pneumoniae* polypeptide. Orthologues of a *Streptococcus pneumoniae* polypeptide can readily be identified as comprising an amino acid sequence that is substantially homologous to one of SEQ ID NO:216 through SEQ ID NO:430 or SEQ ID NO: 592 through SEQ ID NO: 752.

- Modifications and changes can be made in the structure of a polypeptide of the present invention and still obtain a molecule having *Streptococcus pneumoniae* antigenicity. For example, certain amino acids can be substituted for other amino acids in a sequence without appreciable loss of antigenicity. Because it is the interactive capacity and nature of a polypeptide that defines that polypeptide's biological functional activity, certain amino acid sequence substitutions can be made in a polypeptide sequence (or, of course, its underlying DNA coding sequence) and nevertheless obtain a polypeptide with like properties.

- In making such changes, the hydropathic index of amino acids can be considered. The importance of the hydropathic amino acid index in conferring interactive biologic function on a polypeptide is generally understood in the art (Kyte & Doolittle, 1982). It is known that certain amino acids can be substituted for other amino acids having a similar hydropathic index or score and still result in a

polypeptide with similar biological activity. Each amino acid has been assigned a hydropathic index on the basis of its hydrophobicity and charge characteristics. Those indices are: isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (-0.4);  
5 threonine (-0.7); serine (-0.8); tryptophan (-0.9); tyrosine (-1.3); proline (-1.6); histidine (-3.2); glutamate (-3.5); glutamine (-3.5); aspartate (-3.5); asparagine (-3.5); lysine (-3.9); and arginine (-4.5).

It is believed that the relative hydropathic character of the amino acid residue determines the secondary and tertiary structure of the resultant polypeptide, which in  
10 turn defines the interaction of the polypeptide with other molecules, such as enzymes, substrates, receptors, antibodies, antigens, and the like. It is known in the art that an amino acid can be substituted by another amino acid having a similar hydropathic index and still obtain a functionally equivalent polypeptide. In such changes, the substitution of amino acids whose hydropathic indices are within  $\pm 2$  is  
15 preferred, those that are within  $\pm 1$  are particularly preferred, and those within  $\pm 0.5$  are even more particularly preferred.

Substitution of like amino acids can also be made on the basis of hydrophilicity, particularly where the biological functional equivalent polypeptide or peptide thereby created is intended for use in immunological embodiments. U.S.  
20 Patent 4,554,101, incorporated hereinafter by reference, states that the greatest local average hydrophilicity of a polypeptide, as governed by the hydrophilicity of its adjacent amino acids, correlates with its immunogenicity and antigenicity, *i.e.* with a biological property of the polypeptide.

As detailed in U.S. Patent 4,554,101, the following hydrophilicity values have  
25 been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0  $\pm 1$ ); glutamate (+3.0  $\pm 1$ ); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); proline (-0.5  $\pm 1$ ); threonine (-0.4); alanine (-0.5); histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); tryptophan (-3.4). It is understood that an amino acid can be  
30 substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent, and in particular, an immunologically equivalent polypeptide. In such changes, the substitution of amino acids whose hydrophilicity values are

within  $\pm 2$  is preferred, those that are within  $\pm 1$  are particularly preferred, and those within  $\pm 0.5$  are even more particularly preferred.

As outlined above, amino acid substitutions are generally therefore based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like. Exemplary substitutions which take various of the foregoing characteristics into consideration are well known to those of skill in the art and include: arginine and lysine; glutamate and aspartate; serine and threonine; glutamine and asparagine; and valine, leucine and isoleucine (See Table 16, below). The present invention thus contemplates functional or biological equivalents of a *Streptococcus pneumoniae* polypeptide as set forth above.

**TABLE 16**  
**Amino Acid Substitutions**

Original Residue	Exemplary Residue Substitution
Ala	Gly; Ser
Arg	Lys
Asn	Gln; His
Asp	Glu
Cys	Ser
Gln	Asn
Glu	Asp
Gly	Ala
His	Asn; Gln
Ile	Leu; Val
Leu	Ile; Val
Lys	Arg
Met	Leu; Tyr
Ser	Thr
Thr	Ser
Trp	Tyr
Tyr	Trp; Phe
Val	Ile; Leu

15

Biological or functional equivalents of a polypeptide can also be prepared using site-specific mutagenesis. Site-specific mutagenesis is a technique useful in the preparation of second generation polypeptides, or biologically functional equivalent polypeptides or peptides, derived from the sequences thereof, through specific mutagenesis of the underlying DNA. As noted above, such changes can be

20

desirable where amino acid substitutions are desirable. The technique further provides a ready ability to prepare and test sequence variants, for example, incorporating one or more of the foregoing considerations, by introducing one or more nucleotide sequence changes into the DNA. Site-specific mutagenesis allows the production of mutants through the use of specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a primer sequence of sufficient size and sequence complexity to form a stable duplex on both sides of the deletion junction being traversed. Typically, a primer of about 17 to 25 nucleotides in length is preferred, with about 5 to 10 residues on both sides of the junction of the sequence being altered.

In general, the technique of site-specific mutagenesis is well known in the art. As will be appreciated, the technique typically employs a phage vector which can exist in both a single stranded and double stranded form. Typically, site-directed mutagenesis in accordance herewith is performed by first obtaining a single-stranded vector which includes within its sequence a DNA sequence which encodes all or a portion of the *Streptococcus pneumoniae* polypeptide sequence selected. An oligonucleotide primer bearing the desired mutated sequence is prepared (e.g., synthetically). This primer is then annealed to the singled-stranded vector, and extended by the use of enzymes such as *E. coli* polymerase I Klenow fragment, in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed wherein one strand encodes the original non-mutated sequence and the second strand bears the desired mutation. This heteroduplex vector is then used to transform appropriate cells such as *E. coli* cells and clones are selected which include recombinant vectors bearing the mutation. Commercially available kits come with all the reagents necessary, except the oligonucleotide primers.

A *Streptococcus pneumoniae* polypeptide or polypeptide antigen of the present invention is understood to be any *Streptococcus pneumoniae* polypeptide comprising substantial sequence similarity, structural similarity and/or functional similarity to a *Streptococcus pneumoniae* polypeptide comprising the amino acid sequence of one of SEQ ID NO:216 through SEQ ID NO:430 or SEQ ID NO: 592 through SEQ ID NO: 752. In addition, a *Streptococcus pneumoniae* polypeptide or polypeptide antigen of the invention is not limited to a particular source. Thus, the

invention provides for the general detection and isolation of the polypeptides from a variety of sources.

It is contemplated in the present invention, that a *Streptococcus pneumoniae* polypeptide may advantageously be cleaved into fragments for use in further structural or functional analysis, or in the generation of reagents such as *Streptococcus pneumoniae*-related polypeptides and *Streptococcus pneumoniae*-specific antibodies. This can be accomplished by treating purified or unpurified *Streptococcus pneumoniae* polypeptides with a peptidase such as endoproteinase glu-C (Boehringer, Indianapolis, IN). Treatment with CNBr is another method by which peptide fragments may be produced from natural *Streptococcus pneumoniae* polypeptides. Recombinant techniques also can be used to produce specific fragments of a *Streptococcus pneumoniae* polypeptide.

In addition, the inventors also contemplate that compounds sterically similar to a particular *Streptococcus pneumoniae* polypeptide antigen may be formulated to mimic the key portions of the peptide structure, called peptidomimetics. Mimetics are peptide-containing molecules which mimic elements of protein secondary structure. (see, e.g. Johnson *et al.*, 1993). The underlying rationale behind the use of peptide mimetics is that the peptide backbone of proteins exists chiefly to orient amino acid side chains in such a way as to facilitate molecular interactions, such as those of receptor and ligand.

Successful applications of the peptide mimetic concept have thus far focused on mimetics of  $\beta$ -turns within proteins. Likely  $\beta$ -turn structures within *Streptococcus pneumoniae* can be predicted by computer-based algorithms as discussed above. Once the component amino acids of the turn are determined, mimetics can be constructed to achieve a similar spatial orientation of the essential elements of the amino acid side chains, as discussed in Johnson *et al.*, 1993.

Fragments of the *Streptococcus pneumoniae* polypeptides are also included in the invention. A fragment is a polypeptide having an amino acid sequence that entirely is the same as part, but not all, of the amino acid sequence. The fragment can comprise, for example, at least 7 or more (e.g., 8, 10, 12, 14, 16, 18, 20, or more) contiguous amino acids of an amino acid sequence of one of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID NO: 592 through SEQ ID NO: 752. Fragments may be "freestanding" or comprised within a larger polypeptide of which they form a

part or region, most preferably as a single, continuous region. In one embodiment, the fragments include at least one epitope of the mature polypeptide sequence.

- "Fusion protein" refers to a protein or polypeptide encoded by two, often unrelated, fused genes or fragments thereof. For example, fusion proteins or polypeptides comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof have been described. In many cases, employing an immunoglobulin Fc region as a part of a fusion protein or polypeptide is advantageous for use in therapy and diagnosis resulting in, for example, improved pharmacokinetic properties (*see e.g.*, International Application EP-A 0232 2621). On the other hand, for some uses it would be desirable to be able to delete the Fc part after the fusion protein or polypeptide has been expressed, detected and purified.

#### **D. *STREPTOCOCCUS PNEUMONIAE* POLYNUCLEOTIDE AND POLYPEPTIDE VARIANTS**

- "Variant" as the term is used hereinafter, is a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide respectively, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis.

"Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in (Computational Molecular Biology, Lesk, A. M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D. W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A. M., and Griffin, H. G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., SIAM J. Applied Math., 48: 1073 (1988). Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG program package (Devereux, J., *et al* 1984), BLASTP, BLASTN, TBLASTN and FASTA (Altschul, S. F., *et al.*, 1990). The BLASTX program is publicly available from NCBI and other sources (BLAST Manual, Altschul, S., *et al.*, NCBI NLM NIH Bethesda, Md. 20894; Altschul, S., *et al.*, 1990). The well known Smith-Waterman algorithm may also be used to determine identity.

By way of example, a polynucleotide sequence of the present invention may be identical to the reference sequence of one of SEQ ID NO:1 through SEQ ID NO:215 or SEQ ID NO: 431 through SEQ ID NO: 591, that is be 100% identical, or it may include up to a certain integer number of nucleotide alterations as compared to the reference sequence. Such alterations are selected from the group consisting of at least one nucleotide deletion, substitution, including transition and transversion, or insertion, and wherein said alterations may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among the nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence. The number of



nucleotide alterations is determined by multiplying the total number of nucleotides in one of SEQ ID NO:1 through SEQ ID NO:215 or SEQ ID NO: 431 through SEQ ID NO: 591 by the numerical percent of the respective percent identity (divided by 100) and subtracting that product from said total number of nucleotides in one of SEQ ID NO:1 through SEQ ID NO:215 or SEQ ID NO: 431 through SEQ ID NO: 591.

For example, an isolated *Streptococcus pneumoniae* polynucleotide comprising a polynucleotide sequence that has at least 70% identity to the nucleic acid sequence of one of SEQ ID NO:1 through SEQ ID NO:215 or SEQ ID NO: 431 through SEQ ID NO: 591; a degenerate variant thereof or a fragment thereof, wherein the polynucleotide sequence may include up to  $n_n$  nucleic acid alterations over the entire polynucleotide region of the nucleic acid sequence of one of SEQ ID NO:1 through SEQ ID NO:215 or SEQ ID NO: 431 through SEQ ID NO: 591, wherein  $n_n$  is the maximum number of alterations and is calculated by the formula:

$$n_n \leq x_n - (x_n \cdot y),$$

in which  $x_n$  is the total number of nucleic acids of one of SEQ ID NO:1 through SEQ ID NO:215 or SEQ ID NO: 431 through SEQ ID NO: 591 and  $y$  has a value of 0.70, wherein any non-integer product of  $x_n$  and  $y$  is rounded down to the nearest integer prior to subtracting such product from  $x_n$ . Of course,  $y$  may also have a value of 0.80 for 80%, 0.85 for 85%, 0.90 for 90% 0.95 for 95%, etc. Alterations of a polynucleotide sequence encoding one of the polypeptides of SEQ ID NO:216 through SEQ ID NO:430 or SEQ ID NO: 592 through SEQ ID NO: 752 may create nonsense, missense or frameshift mutations in this coding sequence and thereby alter the polypeptide encoded by the polynucleotide following such alterations.

Similarly, a polypeptide sequence of the present invention may be identical to the reference sequence of SEQ ID NO:216 through SEQ ID NO:430 or SEQ ID NO: 592 through SEQ ID NO: 752, that is be 100% identical, or it may include up to a certain integer number of amino acid alterations as compared to the reference sequence such that the % identity is less than 100%. Such alterations are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion, and wherein said alterations may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or anywhere between those terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or

- more contiguous groups within the reference sequence. The number of amino acid alterations for a given % identity is determined by multiplying the total number of amino acids in one of SEQ ID NO:216 through SEQ ID NO:430 or SEQ ID NO: 592 through SEQ ID NO: 752 by the numerical percent of the respective percent identity (divided by 100) and then subtracting that product from said total number of amino acids in one of SEQ ID NO:216 through SEQ ID NO:430 or SEQ ID NO: 592 through SEQ ID NO: 752, or:

$$n_a \leq x_a - (x_a \cdot y),$$

- wherein  $n_a$  is the number of amino acid alterations,  $x_a$  is the total number of amino acids in one of SEQ ID NO:216 through SEQ ID NO:430 SEQ ID NO: 592 through SEQ ID NO: 752, and  $y$  is, for instance 0.70 for 70%, 0.80 for 80%, 0.85 for 85% etc., and wherein any non-integer product of  $x_a$  and  $y$  is rounded down to the nearest integer prior to subtracting it from  $x_a$ .

15

#### **E. VECTORS, HOST CELLS AND RECOMBINANT *STREPTOCOCCUS PNEUMONIAE* POLYPEPTIDES**

- In a preferred embodiment, the present invention provides expression vectors comprising ORF polynucleotides that encode *Streptococcus pneumoniae* polypeptides. Preferably, the expression vectors of the present invention comprise ORF polynucleotides that encode *Streptococcus pneumoniae* polypeptides comprising the amino acid residue sequence of one of SEQ ID NO:216 through SEQ ID NO:430 or SEQ ID NO: 592 through SEQ ID NO: 752. More preferably, the expression vectors of the present invention comprise a polynucleotide comprising the nucleotide base sequence of one of SEQ ID NO:1 through SEQ ID NO:215 or SEQ ID NO: 431 through SEQ ID NO: 591. Even more preferably, the expression vectors of the invention comprise a polynucleotide operatively linked to an enhancer-promoter. More preferably still, the expression vectors of the invention comprise polynucleotide operatively linked to a prokaryotic promoter. Alternatively, the expression vectors of the present invention comprise polynucleotide operatively linked to an enhancer-promoter that is a eukaryotic promoter, and the expression vectors further comprise a polyadenylation signal that is positioned 3' of the carboxy-terminal amino acid and within a transcriptional unit of the encoded polypeptide.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein.

- 5 Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable  
10 separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase.

- Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988), pMAL (New England Biolabs, Beverly, MA) and pRIT5  
15 (Pharmacia, Piscataway, NJ) which fuse glutathione S- transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

- In one embodiment, the coding sequence of the *Streptococcus pneumoniae* polynucleotide is cloned into a pGEX expression vector to create a vector encoding a fusion protein comprising, from the N-terminus to the C-terminus, GST-thrombin  
20 cleavage site-*Streptococcus pneumoniae* polypeptide. The fusion protein can be purified by affinity chromatography using glutathione-agarose resin. Recombinant *Streptococcus pneumoniae* polypeptide unfused to GST can be recovered by cleavage of the fusion protein with thrombin.

- Examples of suitable inducible non-fusion *E. coli* expression vectors include  
25 pTrc (Amann *et al.*, 1988), pET 11d (Studier *et al.*, 1990), pBAD and pCRT7. Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gnl 0-lac fusion promoter mediated by a coexpressed viral RNA polymerase J7 gnl. This viral polymerase is supplied by host  
30 strains BL21 (DE3) or HMS 174(DE3) from a resident prophage harboring a T7 gnl gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacterium with an impaired capacity to proteolytically

cleave the recombinant protein. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli*. Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA mutagenesis or synthesis techniques.

In another embodiment, the *Streptococcus pneumoniae* polynucleotide expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec I (Baldari, *et al.*, 1987), pMFa (Kurjan and Herskowitz, 1982), pJRY88 (Schultz *et al.*, 1987), and pYES2 (Invitrogen Corporation, San Diego, CA).

Alternatively, a *Streptococcus pneumoniae* polynucleotide can be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, Sf 9 cells) include the pAc series (Smith *et al.*, 1983) and the pVL series (Lucklow and Summers, 1989).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987) and pMT2PC (Kaufman *et al.*, 1987). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements.

As used hereinafter, a promoter is a region of a DNA molecule typically within about 100 nucleotide pairs in front of (upstream of) the point at which transcription begins (*i.e.*, a transcription start site). That region typically contains several types of DNA sequence elements that are located in similar relative positions in different genes. As used hereinafter, the term "promoter" includes what is referred to in the art as an upstream promoter region, a promoter region or a promoter of a generalized eukaryotic RNA Polymerase II transcription unit.

Another type of discrete transcription regulatory sequence element is an enhancer. An enhancer provides specificity of time, location and expression level for a particular encoding region (*e.g.*, gene). A major function of an enhancer is to increase the level of transcription of a coding sequence in a cell that contains one or more transcription factors that bind to that enhancer. Unlike a promoter, an enhancer

can function when located at variable distances from transcription start sites so long as a promoter is present.

As used hereinafter, the phrase "enhancer-promoter" means a composite unit that contains both enhancer and promoter elements. An enhancer-promoter is  
5 operatively linked to a coding sequence that encodes at least one gene product. As used hereinafter, the phrase "operatively linked" means that an enhancer-promoter is connected to a coding sequence in such a way that the transcription of that coding sequence is controlled and regulated by that enhancer-promoter. Means for  
10 operatively linking an enhancer-promoter to a coding sequence are well known in the art. As is also well known in the art, the precise orientation and location relative to a coding sequence whose transcription is controlled, is dependent *inter alia* upon the specific nature of the enhancer-promoter. Thus, a TATA box minimal promoter is typically located from about 25 to about 30 base pairs upstream of a transcription  
15 initiation site and an upstream promoter element is typically located from about 100 to about 200 base pairs upstream of a transcription initiation site. In contrast, an enhancer can be located downstream from the initiation site and can be at a considerable distance from that site.

An enhancer-promoter used in a vector construct of the present invention can be any enhancer-promoter that drives expression in a cell to be transfected. By  
20 employing an enhancer-promoter with well-known properties, the level and pattern of gene product expression can be optimized.

For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of  
25 Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual" 2nd, ed, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989, incorporated hereinafter by reference.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell  
30 type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.*, 1987), lymphoid-specific promoters (Calame and Eaton, 1988), in particular,

promoters of T cell receptors (Winoto and Baltimore, 1989) and immunoglobulins (Banerji *et al.*, 1983), Queen and Baltimore (1983), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle, 1989), pancreas-specific promoters (Edlund *et al.*, 1985), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent 4,873,316 and International Application EP 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss, 1990) and the  $\alpha$ -fetoprotein promoter (Campes and Tilghman, 1989).

The invention further provides a recombinant expression vector comprising a DNA molecule encoding a *Streptococcus pneumoniae* polypeptide cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to *Streptococcus pneumoniae* mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types. For instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably hereinafter. It is understood that such terms refer not only to the particular subject cell, but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used hereinafter. A host cell can be any prokaryotic or eukaryotic cell. For example, a *Streptococcus pneumoniae* polypeptide can be expressed in bacterial cells such as *E. coli*, insect cells (such as Sf9, Sf21), yeast or mammalian cells (such as Chinese hamster ovary cells (CHO), VERO, chick embryo

fibroblasts, BHK cells or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA is introduced into prokaryotic or eukaryotic cells *via* conventional transformation, infection or transfection techniques. As used hereinafter, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, ultrasound or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* ("Molecular Cloning: A Laboratory Manual" 2nd, ed, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and other laboratory manuals.

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) a *Streptococcus pneumoniae* polypeptide. Accordingly, the invention further provides methods for producing a *Streptococcus pneumoniae* polypeptide using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding a *Streptococcus pneumoniae* polypeptide has been introduced) in a suitable medium until the *Streptococcus pneumoniae* polypeptide is produced. In another embodiment, the method further comprises isolating the *Streptococcus pneumoniae* polypeptide from the medium or the host cell.

A coding sequence of an expression vector is operatively linked to a transcription termination region. RNA polymerase transcribes an encoding DNA sequence through a site where polyadenylation occurs. Typically, DNA sequences located a few hundred base pairs downstream of the polyadenylation site serve to terminate transcription. Those DNA sequences are referred to hereinafter as transcription-termination regions. Those regions are required for efficient polyadenylation of transcribed messenger RNA (mRNA). Transcription-termination regions are well known in the art. A preferred transcription-termination region used in an adenovirus vector construct of the present invention comprises a polyadenylation signal of SV40 or the protamine gene.

An expression vector comprises a polynucleotide that encodes a *Streptococcus pneumoniae* polypeptide. Such a polypeptide is meant to include a sequence of nucleotide bases encoding a *Streptococcus pneumoniae* polypeptide sufficient in length to distinguish the segment from a polynucleotide segment encoding a non-*Streptococcus pneumoniae* polypeptide. A polypeptide of the invention can also encode biologically functional polypeptides or peptides which have variant amino acid sequences, such as with changes selected based on considerations such as the relative hydropathic score of the amino acids being exchanged. These variant sequences are those isolated from natural sources or induced in the sequences disclosed hereinafter using a mutagenic procedure such as site-directed mutagenesis.

Preferably, the expression vectors of the present invention comprise polynucleotide that encode polypeptides comprising the amino acid residue sequence of one of SEQ ID NO:216 through SEQ ID NO:430 or SEQ ID NO: 592 through SEQ ID NO: 752. An expression vector can include a *Streptococcus pneumoniae* polypeptide coding region itself of any of the *Streptococcus pneumoniae* polypeptides noted above or it can contain coding regions bearing selected alterations or modifications in the basic coding region of such a *Streptococcus pneumoniae* polypeptide. Alternatively, such vectors or fragments can code larger polypeptides or polypeptides which nevertheless include the basic coding region. In any event, it should be appreciated that due to codon redundancy as well as biological functional equivalence, this aspect of the invention is not limited to the particular DNA molecules corresponding to the polypeptide sequences noted above.

Exemplary vectors include the mammalian expression vectors of the pCMV family including pCMV6b and pCMV6c (Chiron Corp., Emeryville CA.). In certain cases, and specifically in the case of these individual mammalian expression vectors, the resulting constructs can require co-transfection with a vector containing a selectable marker such as pSV2neo. Via co-transfection into a dihydrofolate reductase-deficient Chinese hamster ovary cell line, such as DG44, clones expressing *Streptococcus pneumoniae* polypeptides by virtue of DNA incorporated into such expression vectors can be detected.

A DNA molecule of the present invention can be incorporated into a vector by a number of techniques that are well known in the art. For instance, the vector



pUC18 has been demonstrated to be of particular value in cloning and expression of genes. Likewise, the related vectors M13mp18 and M13mp19 can be used in certain embodiments of the invention, in particular, in performing dideoxy sequencing.

5 An expression vector of the present invention is useful both as a means for preparing quantities of the *Streptococcus pneumoniae* polypeptide-encoding DNA itself, and as a means for preparing the encoded polypeptide and peptides. It is contemplated that where *Streptococcus pneumoniae* polypeptides of the invention are made by recombinant means, one can employ either prokaryotic or eukaryotic expression vectors as shuttle systems.

10 In another aspect, the recombinant host cells of the present invention are prokaryotic host cells. Preferably, the recombinant host cells of the invention are bacterial cells of the DH5  $\alpha$  strain of *Escherichia coli*. In general, prokaryotes are preferred for the initial cloning of DNA sequences and constructing the vectors useful in the invention. For example, *E. coli* K12 strains can be particularly useful. Other  
15 microbial strains that can be used include *E. coli* B, and *E. coli* X1976 (ATCC No. 31537). These examples are, of course, intended to be illustrative rather than limiting.

The aforementioned strains, as well as *E. coli* W3110 (ATCC No. 273325), *E. coli* BL21(DE3), *E. coli* Top10, *bacilli* such as *Bacillus subtilis*, or other  
20 enterobacteriaceae such as *Salmonella typhimurium* (or other attenuated *Salmonella* strains as described in U.S. Patent 4,837,151) or *Serratia marcesans*, and various *Pseudomonas* species can be used.

In general, plasmid vectors containing replicon and control sequences, which are derived from species compatible with the host cell are used in connection with  
25 these hosts. The vector ordinarily carries a replication site, as well as marking sequences which are capable of providing phenotypic selection in transformed cells. For example, *E. coli* can be transformed using pBR322, a plasmid derived from an *E. coli* species (Bolivar, *et al.* 1977). pBR322 contains genes for ampicillin and tetracycline resistance and thus provides easy means for identifying transformed  
30 cells. The pBR plasmid, or other microbial plasmid or phage must also contain, or be modified to contain, promoters which can be used by the microbial organism for expression of its own polypeptides.

Those promoters most commonly used in recombinant DNA construction include the  $\beta$ -lactamase (penicillinase) and lactose promoter systems (Chang, *et al.* 1978; Itakura, *et al.* 1977; Goeddel, *et al.* 1979; Goeddel, *et al.* 1980) and a tryptophan (TRP) promoter system (EP 0036776; Siebwenlist *et al.* 1980). While  
5 these are the most commonly used, other microbial promoters have been discovered and utilized, and details concerning their nucleotide sequences have been published, enabling a skilled worker to introduce functional promoters into plasmid vectors (Siebwenlist, *et al.* 1980).

In addition to prokaryotes, eukaryotic microbes such as yeast can also be  
10 used. *Saccharomyces cerevisiae* or common baker's yeast is the most commonly used among eukaryotic microorganisms, although a number of other strains are commonly available. For expression in *Saccharomyces*, the plasmid YRp7, for example, is commonly used (Stinchcomb, *et al.* 1979; Kingsman, *et al.* 1979; Tschemper, *et al.* 1980). This plasmid already contains the *trp1* gene which provides  
15 a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, for example ATCC No. 44076 or PEP4-1 (Jones, 1977). The presence of the *trp1* lesion as a characteristic of the yeast host cell genome then provides an effective environment for detecting transformation by growth in the absence of tryptophan.

Suitable promoter sequences in yeast vectors include the promoters for 3-phosphoglycerate kinase (Hitzeman, *et al.* 1980) or other glycolytic enzymes (Hess, *et al.* 1968; Holland, *et al.* 1978) such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-  
20 6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase. In constructing suitable expression plasmids, the termination sequences associated with these genes are also introduced into the expression vector downstream from the sequences to be expressed to provide polyadenylation of the mRNA and termination. Other promoters, which have the additional advantage of transcription controlled by  
25 growth conditions are the promoter region for alcohol dehydrogenase 2, isocytichrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, and the aforementioned glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Any plasmid vector

containing a yeast-compatible promoter, origin or replication and termination sequences are suitable.

In addition to microorganisms, cultures of cells derived from multicellular organisms can also be used as hosts. In principle, any such cell culture is workable, whether from vertebrate or invertebrate culture. However, interest has been greatest in vertebrate cells, and propagation of vertebrate cells in culture (tissue culture) has become a routine procedure in recent years. Examples of such useful host cell lines are AtT-20, VERO, HeLa, NSO, PER C6, Chinese hamster ovary (CHO) cell lines, and W138, BHK, COSM6, COS-7, 293 and MDCK cell lines. Expression vectors for such cells ordinarily include (if necessary) an origin of replication, a promoter located upstream of the gene to be expressed, along with any necessary ribosome binding sites, RNA splice sites, polyadenylation site, and transcriptional terminator sequences.

Where expression of recombinant *Streptococcus pneumoniae* polypeptides is desired and a eukaryotic host is contemplated, it is most desirable to employ a vector such as a plasmid, that incorporates a eukaryotic origin of replication. Additionally, for the purposes of expression in eukaryotic systems, one desires to position the *Streptococcus pneumoniae* encoding sequence adjacent to and under the control of an effective eukaryotic promoter such as promoters used in combination with Chinese hamster ovary cells. To bring a coding sequence under control of a promoter, whether it is eukaryotic or prokaryotic, the 5' end of the translation initiation region of the proper translational reading frame of the polypeptide must be positioned between about 1 and about 50 nucleotides 3' of or downstream with respect to the promoter chosen. Furthermore, where eukaryotic expression is anticipated, one would typically desire to incorporate into the transcriptional unit which includes the *Streptococcus pneumoniae* polypeptide.

Means of transforming or transfecting cells with exogenous polynucleotide such as DNA molecules are well known in the art and include techniques such as calcium-phosphate- or DEAE-dextran-mediated transfection, protoplast fusion, electroporation, liposome mediated transfection, direct microinjection and adenovirus infection (see e.g., Sambrook, Fritsch and Maniatis, 1989).

The most widely used method is transfection mediated by either calcium phosphate or DEAE-dextran. Although the mechanism remains obscure, it is

believed that the transfected DNA enters the cytoplasm of the cell by endocytosis and is transported to the nucleus. Depending on the cell type, up to 90% of a population of cultured cells can be transfected at any one time. Because of its high efficiency, transfection mediated by calcium phosphate or DEAE-dextran is the method of choice for experiments that require transient expression of the foreign DNA in large numbers of cells. Calcium phosphate-mediated transfection is also used to establish cell lines that integrate copies of the foreign DNA, which are usually arranged in head-to-tail tandem arrays into the host cell genome.

In the protoplast fusion method, protoplasts derived from bacteria carrying high numbers of copies of a plasmid of interest are mixed directly with cultured mammalian cells. After fusion of the cell membranes (usually with polyethylene glycol), the contents of the bacteria are delivered into the cytoplasm of the mammalian cells and the plasmid DNA is transported to the nucleus. Protoplast fusion is not as efficient as transfection for many of the cell lines that are commonly used for transient expression assays, but it is useful for cell lines in which endocytosis of DNA occurs inefficiently. Protoplast fusion frequently yields multiple copies of the plasmid DNA tandemly integrated into the host chromosome.

The application of brief, high-voltage electric pulses to a variety of mammalian and plant cells leads to the formation of nanometer-sized pores in the plasma membrane. DNA is taken directly into the cell cytoplasm either through these pores or as a consequence of the redistribution of membrane components that accompanies closure of the pores. Electroporation can be extremely efficient and can be used both for transient expression of cloned genes and for establishment of cell lines that carry integrated copies of the gene of interest. Electroporation, in contrast to calcium phosphate-mediated transfection and protoplast fusion, frequently gives rise to cell lines that carry one, or at most a few, integrated copies of the foreign DNA.

Liposome transfection involves encapsulation of DNA and RNA within liposomes, followed by fusion of the liposomes with the cell membrane. The mechanism of how DNA is delivered into the cell is unclear but transfection efficiencies can be as high as 90%.

Direct microinjection of a DNA molecule into nuclei has the advantage of not exposing DNA to cellular compartments such as low-pH endosomes. Microinjection

is therefore used primarily as a method to establish lines of cells that carry integrated copies of the DNA of interest.

The use of adenovirus as a vector for cell transfection is well known in the art. Adenovirus vector-mediated cell transfection has been reported for various cells (Stratford-Perricaudet, *et al.* 1992).

A transfected cell can be prokaryotic or eukaryotic. Preferably, the host cells of the invention are prokaryotic host cells. Where it is of interest to produce a *Streptococcus pneumoniae* polypeptide, cultured prokaryotic host cells are of particular interest.

In yet another embodiment, the present invention contemplates a process or method of preparing *Streptococcus pneumoniae* polypeptides comprising transforming, transfecting or infecting cells with a polynucleotide that encodes a *Streptococcus pneumoniae* polypeptide to produce transformed host cells; and maintaining the transformed host cells under biological conditions sufficient for expression of the polypeptide. Preferably, the transformed host cells are prokaryotic cells. Alternatively, the host cells are eukaryotic cells. More preferably, the prokaryotic cells are bacterial cells of the DH5- $\alpha$  strain of *Escherichia coli*. Even more preferably, the polynucleotide transfected into the transformed cells comprise the nucleic acid sequence of one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO: 431 through SEQ ID NO: 591. Additionally, transfection is accomplished using an expression vector disclosed above. A host cell used in the process is capable of expressing a functional, recombinant *Streptococcus pneumoniae* polypeptide.

Following transfection, the cell is maintained under culture conditions for a period of time sufficient for expression of a *Streptococcus pneumoniae* polypeptide. Culture conditions are well known in the art and include ionic composition and concentration, temperature, pH and the like. Typically, transfected cells are maintained under culture conditions in a culture medium. Suitable media for various cell types are well known in the art. In a preferred embodiment, temperature is from about 20°C to about 50°C, more preferably from about 30°C to about 40°C and, even more preferably about 37°C.

The pH is preferably from about a value of 6.0 to a value of about 8.0, more preferably from about a value of about 6.8 to a value of about 7.8 and, most preferably about 7.4. Osmolality is preferably from about 200 milliosmols per liter

(mosm/L) to about 400 mosm/l and, more preferably from about 290 mosm/L to about 310 mosm/L. Other biological conditions needed for transfection and expression of an encoded protein are well known in the art.

- 5 Transfected cells are maintained for a period of time sufficient for expression of an *Streptococcus pneumoniae* polypeptide. A suitable time depends *inter alia* upon the cell type used and is readily determinable by a skilled artisan. Typically, maintenance time is from about 2 to about 14 days.

- 10 Recombinant *Streptococcus pneumoniae* polypeptide is recovered or collected either from the transfected cells or the medium in which those cells are cultured. Recovery comprises isolating and purifying the *Streptococcus pneumoniae* polypeptide. Isolation and purification techniques for polypeptides are well known in the art and include such procedures as precipitation, filtration, chromatography, electrophoresis and the like.

15 **F. ANTIBODIES IMMUNOREACTIVE WITH STREPTOCOCCUS PNEUMONIAE POLYPEPTIDES**

- In still another embodiment, the present invention provides antibodies immunoreactive with *Streptococcus pneumoniae* polypeptides. Preferably, the antibodies of the invention are monoclonal antibodies. Additionally, the  
20 *Streptococcus pneumoniae* polypeptides comprise the amino acid residue sequence of one of SEQ ID NO:216 through SEQ ID NO:430 or SEQ ID NO: 592 through SEQ ID NO: 752. Means for preparing and characterizing antibodies are well known in the art (See, e.g., Antibodies "A Laboratory Manual", E. Harlow and D. Lane, Cold Spring Harbor Laboratory, 1988).

- 25 Briefly, a polyclonal antibody is prepared by immunizing an animal with an immunogen comprising a polypeptide or polynucleotide of the present invention, and collecting antisera from that immunized animal. A wide range of animal species can be used for the production of antisera. Typically an animal used for production of anti-antisera is a rabbit, a mouse, a rat, a hamster or a guinea pig. Because of the  
30 relatively large blood volume of rabbits, a rabbit is a preferred choice for production of polyclonal antibodies.

As is well known in the art, a given polypeptide or polynucleotide may vary in its immunogenicity. It is often necessary therefore to couple the immunogen (e.g., a

polypeptide or polynucleotide) of the present invention with a carrier. Exemplary and preferred carriers are CRM<sub>197</sub>, keyhole limpet hemocyanin (KLH) and bovine serum albumin (BSA). Other albumins such as ovalbumin, mouse serum albumin or rabbit serum albumin can also be used as carriers.

- 5 Means for conjugating a polypeptide or a polynucleotide to a carrier protein are well known in the art and include glutaraldehyde, m-maleimidobencoyl-N-hydroxysuccinimide ester, carbodiimide and bis-biazotized benzidine.

- The amount of immunogen used for the production of polyclonal antibodies varies *inter alia*, upon the nature of the immunogen as well as the animal used for  
10 immunization. A variety of routes can be used to administer the immunogen (subcutaneous, intramuscular, intradermal, intravenous and intraperitoneal). The production of polyclonal antibodies is monitored by sampling blood of the immunized animal at various points following immunization. When a desired level of immunogenicity is obtained, the immunized animal can be bled and the serum  
15 isolated and stored.

- In another aspect, the present invention contemplates a process of producing an antibody immunoreactive with a *Streptococcus pneumoniae* polypeptide comprising the steps of (a) transfecting recombinant host cells with a polynucleotide that encodes a *Streptococcus pneumoniae* polypeptide; (b) culturing the host cells  
20 under conditions sufficient for expression of the polypeptide; (c) recovering the polypeptides; and (d) preparing the antibodies to the polypeptides. Preferably, the host cell is transfected with the polynucleotide of one of SEQ ID NO:1 through SEQ ID NO:215 or SEQ ID NO: 431 through SEQ ID NO: 591. Even more preferably, the present invention provides antibodies prepared according to the process described  
25 above.

- A monoclonal antibody of the present invention can be readily prepared through use of well-known techniques such as those exemplified in U.S. Patent 4,196,265, hereinafter incorporated by reference. Typically, a technique involves first  
30 immunizing a suitable animal with a selected antigen (e.g., a polypeptide or polynucleotide of the present invention) in a manner sufficient to provide an immune response. Rodents, such as mice and rats, are preferred animals. Spleen cells from the immunized animal are then fused with cells of an immortal myeloma cell. Where

the immunized animal is a mouse, a preferred myeloma cell is a murine NS-1 myeloma cell.

The fused spleen/myeloma cells are cultured in a selective medium to select fused spleen/myeloma cells from the parental cells. Fused cells are separated from the mixture of non-fused parental cells, e.g., by the addition of agents that block the *de novo* synthesis of nucleotides in the tissue culture media. Exemplary and preferred agents are aminopterin, methotrexate, and azaserine. Aminopterin and methotrexate block *de novo* synthesis of both purines and pyrimidines, whereas azaserine blocks only purine synthesis. Where aminopterin or methotrexate is used, the media is supplemented with hypoxanthine and thymidine as a source of nucleotides. Where azaserine is used, the media is supplemented with hypoxanthine.

This culturing provides a population of hybridomas from which specific hybridomas are selected. Typically, selection of hybridomas is performed by culturing the cells by single-clone dilution in microtiter plates, followed by testing the individual clonal supernatants for reactivity with an antigen-polypeptide. The selected clones can then be propagated indefinitely to provide the monoclonal antibody.

By way of specific example, to produce an antibody of the present invention, mice are injected intraperitoneally with between about 1-200  $\mu\text{g}$  of an antigen comprising a polypeptide of the present invention. B lymphocyte cells are stimulated to grow by injecting the antigen in association with an adjuvant such as complete Freund's adjuvant (a non-specific stimulator of the immune response containing killed *Mycobacterium tuberculosis*). At some time (e.g., at least two weeks) after the first injection, mice are boosted by injection with a second dose of the antigen mixed with incomplete Freund's adjuvant.

A few weeks after the second injection, mice are tail bled and the sera titered by immunoprecipitation against radiolabeled antigen. Preferably, the process of boosting and titering is repeated until a suitable titer is achieved. The spleen of the mouse with the highest titer is removed and the spleen lymphocytes are obtained by homogenizing the spleen with a syringe. Typically, a spleen from an immunized mouse contains approximately  $5 \times 10^7$  to  $2 \times 10^8$  lymphocytes.



5 Mutant lymphocyte cells known as myeloma cells are obtained from laboratory animals in which such cells have been induced to grow by a variety of well-known methods. Myeloma cells lack the salvage pathway of nucleotide biosynthesis. Because myeloma cells are tumor cells, they can be propagated indefinitely in tissue culture, and are thus denominated immortal. Numerous cultured cell lines of myeloma cells from mice and rats, such as murine NS-1 myeloma cells, have been established.

10 Myeloma cells are combined under conditions appropriate to foster fusion with the normal antibody-producing cells from the spleen of the mouse or rat injected with the antigen/polypeptide of the present invention. Fusion conditions include, for example, the presence of polyethylene glycol. The resulting fused cells are hybridoma cells. Like myeloma cells, hybridoma cells grow indefinitely in culture.

15 Hybridoma cells are separated from unfused myeloma cells by culturing in a selection medium such as HAT media (hypoxanthine, aminopterin, thymidine). Unfused myeloma cells lack the enzymes necessary to synthesize nucleotides from the salvage pathway because they are killed in the presence of aminopterin, methotrexate, or azaserine. Unfused lymphocytes also do not continue to grow in tissue culture. Thus, only cells that have successfully fused (hybridoma cells) can grow in the selection media.

20 Each of the surviving hybridoma cells produces a single antibody. These cells are then screened for the production of the specific antibody immunoreactive with an antigen/polypeptide of the present invention. Single cell hybridomas are isolated by limiting dilutions of the hybridomas. The hybridomas are serially diluted many times and, after the dilutions are allowed to grow, the supernatant is tested for the presence of the monoclonal antibody. The clones producing that antibody are then cultured in large amounts to produce an antibody of the present invention in convenient quantity.

25 By use of a monoclonal antibody of the present invention, specific polypeptides and polynucleotide of the invention are identified as antigens. Once identified, those polypeptides and polynucleotides are isolated and purified by techniques such as antibody-affinity chromatography. In antibody-affinity chromatography, a monoclonal antibody is bound to a solid substrate and exposed to a solution containing the desired antigen. The antigen is removed from the solution

through an immunospecific reaction with the bound antibody. The polypeptide or polynucleotide is then easily removed from the substrate and purified.

Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent 5,223,409; International Application WO 92/18619; International Application WO 91/17271; International Application WO 92/20791; International Application WO 92/15679; International Application WO 93/01288; International Application WO 92/01047; International Application WO 92/09690; International Application WO 90/02809.

Additionally, recombinant anti-*Streptococcus pneumoniae* antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human fragments, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in International Application PCT/US86/02269; International Application EP 184,187; International Application EP 171,496; International Application EP 173,494; International Application WO 86/01533; U.S. Patent 4,816,567; and International Application EP 125,023.

An anti-*Streptococcus pneumoniae* antibody (e.g., monoclonal antibody) is used to isolate *Streptococcus pneumoniae* polypeptides by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-*Streptococcus pneumoniae* antibody facilitates the purification of a natural *Streptococcus pneumoniae* polypeptide from cells and recombinantly produced *Streptococcus pneumoniae* polypeptides expressed in host cells. Moreover, an anti-*Streptococcus pneumoniae* antibody is used to detect *Streptococcus pneumoniae* polypeptide (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance of the *Streptococcus pneumoniae* polypeptide. The detection of circulating fragments of a *Streptococcus pneumoniae* polypeptide is used to identify *Streptococcus pneumoniae* polypeptide turnover in a subject. Anti-*Streptococcus pneumoniae* antibodies are used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection is facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include

various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, P-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and acquerin, and examples of suitable radioactive material include  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{15}\text{S}$  or  $^3\text{H}$ .

#### G. PHARMACEUTICAL AND IMMUNOGENIC COMPOSITIONS

In certain embodiments, the present invention provides pharmaceutical and immunogenic compositions comprising *Streptococcus pneumoniae* polypeptides and physiologically acceptable carriers. More preferably, the pharmaceutical compositions comprise one or more *Streptococcus pneumoniae* polypeptides comprising the amino acid residue sequence of one or more of SEQ ID NO:216 through SEQ ID NO:430 or SEQ ID NO: 592 through SEQ ID NO: 752. In other embodiments, the pharmaceutical compositions of the invention comprise polynucleotides that encode *Streptococcus pneumoniae* polypeptides, and physiologically acceptable carriers. Preferably, the pharmaceutical and immunogenic compositions of the present invention comprise *Streptococcus pneumoniae* polypeptides comprising the amino acid sequence of one of SEQ ID NO:216 through SEQ ID NO:430 or SEQ ID NO: 592 through SEQ ID NO: 752. Alternatively, the pharmaceutical and immunogenic compositions comprise polynucleotides comprising the nucleotide sequence of one of SEQ ID NO:1 through SEQ ID NO:215 or SEQ ID NO: 431 through SEQ ID NO: 591.

Various tests are used to assess the *in vitro* immunogenicity of the polypeptides of the invention. For example, an *in vitro* opsonic assay is conducted by incubating together a mixture of *Streptococcus pneumoniae* cells, heat inactivated human serum containing specific antibodies to the polypeptide in question, and an exogenous complement source. Opsonophagocytosis proceeds during incubation of freshly isolated human polymorphonuclear cells (PMN's) and the

antibody/complement/pneumococcal cell mixture. Bacterial cells that are coated with antibody and complement are killed upon opsonophagocytosis. Colony forming units (cfu) of surviving bacteria that escape from opsonophagocytosis are determined by plating the assay mixture. Titers are reported as the reciprocal of the highest dilution that gives  $\geq 50\%$  bacterial killing, as determined by comparison to assay controls. Specimens which demonstrate less than 50% killing at the lowest serum dilution tested (1:8), are reported as having an OPA titer of 4. The highest dilution tested is 1:2560. Samples with  $\geq 50\%$  killing at the highest dilution are repeated, beginning with a higher initial dilution. The method described above is a modification of Gray's method (Gray, 1990).

A test serum control, which contains test serum plus bacterial cells and heat inactivated complement, is included for each individual serum. This control can be used to assess whether the presence of antibiotics or other serum components are capable of killing the bacterial strain directly (*i.e.* in the absence of complement or PMN's). A human serum with known opsonic titer is used as a positive human serum control. The opsonic antibody titer for each unknown serum can be calculated as the reciprocal of the initial dilution of serum giving 50% cfu reduction compared to the control without serum.

A whole cell ELISA assay is also used to assess *in vitro* immunogenicity and surface exposure of the polypeptide antigen, wherein the bacterial strain of interest (*S. pneumoniae*) is coated onto a plate, such as a 96 well plate, and test sera from an immunized animal is reacted with the bacterial cells. If any antibody, specific for the test polypeptide antigen, is reactive with a surface exposed epitope of the polypeptide antigen, it can be detected by standard methods known to one skilled in the art.

Any polypeptide demonstrating the desired *in vitro* activity is then tested in an *in vivo* animal challenge model. In certain embodiments, immunogenic compositions are used in the immunization of an animal (*e.g.*, a mouse) by methods and routes of immunization known to those of skill in the art (*e.g.*, intranasal, parenteral, oral, rectal, vaginal, transdermal, intraperitoneal, intravenous, subcutaneous, *etc.*). Following immunization of the animal with a particular *Streptococcus pneumoniae* immunogenic composition, the animal is challenged with *Streptococcus pneumoniae* and assayed for resistance to *Streptococcus pneumoniae* infection.

In one embodiment, six-week old, pathogen-free, Balb/c mice are immunized and challenged with *Streptococcus pneumoniae*. For example, BALB/C mice, at 10 animals per group, are immunized (by slow instillation into the nostrils of each mouse) with one or more doses of the desired polypeptide in an immunogenic composition. *Streptococcus pneumoniae* colonizes the nasopharynx of Balb/c mice, but does not cause disease or death. Subsequently, the Balb/c mice are challenged with streptomycin-resistant *Streptococcus pneumoniae*. The Balb/c mice are sacrificed post-challenge, the noses removed, and homogenized in sterile saline. The homogenate is diluted in saline and plated on streptomycin-containing TSA plates. Plates are incubated overnight at 37°C and then colonies are counted. Statistically significant reduction of nasopharyngeal colonization indicates that the polypeptide is suitable for use in human clinical trials.

In another embodiment, six-week old, pathogen-free, male CBA/CaHn xid/J (CBA/N) mice are immunized intranasally or parenterally prior to *Streptococcus pneumoniae* challenge. CBA/N mice, at 10 animals per group, are immunized with an appropriate amount of the desired polypeptide in an immunogenic composition to be tested. CBA/N mice are immunodeficient (XID) and, when challenged with appropriate *Streptococcus pneumoniae*, develop nasopharyngeal colonization, bacteremia and death.

The CBA/N mice are immunized intranasally or subcutaneously with one or more doses of the desired immunogenic composition. Subsequently, the CBA/N mice are challenged with streptomycin-resistant *Streptococcus pneumoniae*. To determine the effects of immunization on intranasal colonization, the CBA/N mice are sacrificed post-challenge, the noses are removed, and homogenized in sterile saline. The homogenate is serially diluted in saline and plated on streptomycin-containing TSA plates. In addition, blood collected post-challenge from each mouse is also plated on streptomycin-containing TSA plates to determine levels of bacteremia. Plates are incubated overnight at 37°C and then colonies are counted. In another embodiment, CBA/N mice are immunized as described above and challenged intranasally. The CBA/N mice are observed daily after challenge, and the mortality is monitored for 14 days. Statistically significant reduction of nasopharyngeal colonization and/or mortality indicates that the polypeptide is suitable for use in human clinical trials.

- The *Streptococcus pneumoniae* polynucleotides, polypeptides, modulators of a *Streptococcus pneumoniae* polypeptides, and anti-*Streptococcus pneumoniae* antibodies (also referred to hereinafter as "active compounds") of the invention are incorporated into pharmaceutical and immunogenic compositions suitable for
- 5 administration to a subject, e.g., a human. Such compositions typically comprise the nucleic acid molecule, protein, modulator, or antibody and a pharmaceutically acceptable carrier. As used hereinafter the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the
- 10 like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, such media can be used in the compositions of the invention. Supplementary active compounds can also be incorporated into the compositions.
- 15 A pharmaceutical or immunogenic composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral (e.g., intravenous, intradermal, subcutaneous, intraperitoneal), transmucosal (e.g., oral, rectal, intranasal, vaginal, respiratory) and transdermal (topical). Solutions or suspensions used for parenteral, intradermal, or
- 20 subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates,
- 25 citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.
- 30 Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor ELTM(BASF, Parsippany, NJ) or phosphate

buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a *Streptococcus pneumoniae* polypeptide or anti-*Streptococcus pneumoniae* antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or

swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer. Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems.

Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods



known to those skilled in the art, for example, as described in U.S. Patent 4,522,811 which is incorporated hereinafter by reference.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit  
5 form as used hereinafter refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the  
10 active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

Combination immunogenic compositions are provided by including two or more of the polypeptides of the invention, as well as by combining one or more of the  
15 polypeptides of the invention with one or more known *S. pyogenes* polypeptides, including, but not limited to, the C5a peptidase, the M proteins, adhesins and the like.

In other embodiments, combination immunogenic compositions are provided by combining one or more of the polypeptides of the invention with one or more known *S. pneumoniae* polysaccharides or polysaccharide-protein conjugates,  
20 including, but not limited to, the currently available 23-valent pneumococcal capsular polysaccharide vaccine and the 7-valent pneumococcal polysaccharide-protein conjugate vaccine.

The nucleic acid molecules of the invention are inserted into a variety of vectors and expression systems. A great variety of expression systems are used.  
25 Such systems include, among others, chromosomal, episomal and virus-derived systems, e.g., vectors derived from bacterial plasmids, attenuated bacteria such as *Salmonella* (U.S. Patent 4,837,151) from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as vaccinia and other poxviruses, sindbis, adenovirus, baculoviruses,  
30 papova viruses, such as SV40, fowl pox viruses, pseudorabies viruses and retroviruses, alphaviruses such as Venezuelan equine encephalitis virus (U.S. Patent 5,643,576); nonsegmented negative-stranded RNA viruses such as vesicular stomatitis virus (U.S. Patent 6,168,943), and vectors derived from combinations

thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression systems should include control regions that regulate as well as engender expression, such as promoters and other regulatory elements (such as a polyadenylation signal). Generally, any system or  
5 vector suitable to maintain, propagate or express polynucleotides to produce a polypeptide in a host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual" 2nd, ed, Cold Spring Harbor Laboratory, Cold Spring  
10 Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

A pharmaceutically acceptable vehicle is understood to designate a compound or a combination of compounds entering into a pharmaceutical or immunogenic composition which does not cause side effects and which makes it possible, for example, to facilitate the administration of the active compound, to  
15 increase its life and/or its efficacy in the body, to increase its solubility in solution or alternatively to enhance its preservation. These pharmaceutically acceptable vehicles are well known and will be adapted by persons skilled in the art according to the nature and the mode of administration of the active compound chosen.

As defined hereinafter, an "adjuvant" is a substance that serves to enhance  
20 the immunogenicity of an "antigen" or the immunogenic compositions comprising a polypeptide antigens having an amino acid sequence chosen from one of SEQ ID NO:216 through SEQ ID NO:430 or SEQ ID NO: 592 through SEQ ID NO: 752. Thus, adjuvants are often given to boost the immune response and are well known to the skilled artisan. Examples of adjuvants contemplated in the present invention  
25 include, but are not limited to, aluminum salts (alum) such as aluminum phosphate and aluminum hydroxide, *Mycobacterium tuberculosis*, *Bordetella pertussis*, bacterial lipopolysaccharides, aminoalkyl glucosamine phosphate compounds (AGP), or derivatives or analogs thereof, which are available from Corixa (Hamilton, MT), and which are described in United States Patent Number 6,113,918; one such AGP is 2-  
30 [(R)-3-tetradecanoyloxytetradecanoylamino]ethyl 2-Deoxy-4-O-phosphono-3-O-[(R)-3-tetradecanoyloxytetradecanoyl]-2-[(R)-3-tetradecanoyloxytetradecanoylamino]-b-D-glucopyranoside, which is also known as 529 (formerly known as RC529), which is formulated as an aqueous form or as a stable emulsion, MPL™ (3-O-deacylated

monophosphoryl lipid A) (Corixa) described in U.S. Patent Number 4,912,094, synthetic polynucleotides such as oligonucleotides containing a CpG motif (U.S. Patent Number 6,207,646), polypeptides, saponins such as Quil A or STIMULON™ QS-21 (Antigenics, Framingham, Massachusetts), described in U.S. Patent Number 5,057,540, a pertussis toxin (PT), or an *E. coli* heat-labile toxin (LT), particularly LT-K63, LT-R72, CT-S109, PT-K9/G129; *see, e.g.*, International Patent Publication Nos. WO 93/13302 and WO 92/19265, cholera toxin (either in a wild-type or mutant form, *e.g.*, wherein the glutamic acid at amino acid position 29 is replaced by another amino acid, preferably a histidine, in accordance with published International Patent Application number WO 00/18434). Various cytokines and lymphokines are suitable for use as adjuvants. One such adjuvant is granulocyte-macrophage colony stimulating factor (GM-CSF), which has a nucleotide sequence as described in U.S. Patent Number 5,078,996. A plasmid containing GM-CSF cDNA has been transformed into *E. coli* and has been deposited with the American Type Culture Collection (ATCC), 1081 University Boulevard, Manassas, VA 20110-2209, under Accession Number 39900. The cytokine Interleukin-12(IL-12) is another adjuvant which is described in U.S. Patent Number 5,723,127. Other cytokines or lymphokines have been shown to have immune modulating activity, including, but not limited to, the interleukins 1-alpha, 1-beta, 2, 4, 5, 6, 7, 8, 10, 13, 14, 15, 16, 17 and 18, the interferons-alpha, beta and gamma, granulocyte colony stimulating factor, and the tumor necrosis factors alpha and beta, and are suitable for use as adjuvants.

A composition of the present invention is typically administered parenterally in dosage unit formulations containing standard, well-known nontoxic physiologically acceptable carriers, adjuvants, and vehicles as desired. The term parenteral as used hereinafter includes intravenous, intra-muscular, intraarterial injection, or infusion techniques.

Injectable preparations, for example sterile injectable aqueous or oleaginous suspensions, are formulated according to the known art using suitable dispersing or wetting agents and suspending agents. The sterile injectable preparation can also be a sterile injectable solution or suspension in a nontoxic parenterally acceptable diluent or solvent, for example, as a solution in 1,3-butanediol.

Among the acceptable vehicles and solvents that may be employed are water, Ringer's solution, and isotonic sodium chloride solution. In addition, sterile,

fixed oils are conventionally employed as a solvent or suspending medium. For this purpose any bland fixed oil can be employed including synthetic mono- or di-glycerides. In addition, fatty acids such as oleic acid find use in the preparation of injectables.

- 5 Preferred carriers include neutral saline solutions buffered with phosphate, lactate, Tris, and the like. Of course, when administering viral vectors, one purifies the vector sufficiently to render it essentially free of undesirable contaminants, such as defective interfering adenovirus particles or endotoxins and other pyrogens such that it does not cause any untoward reactions in the individual receiving the vector
- 10 construct. A preferred means of purifying the vector involves the use of buoyant density gradients, such as cesium chloride gradient centrifugation.

A carrier can also be a liposome. Means for using liposomes as delivery vehicles are well known in the art (see, e.g. Gabizon *et al.*, 1990; Ferruti *et al.*, 1988; and Ranade, 1989).

- 15 The immunogenic compositions of this invention also comprise a polynucleotide sequence of this invention operatively associated with a regulatory sequence that controls gene expression. The polynucleotide sequence of interest is engineered into an expression vector, such as a plasmid, under the control of regulatory elements which will promote expression of the DNA, that is, promoter and/or enhancer elements. In a preferred embodiment, the human cytomegalovirus
- 20 immediate-early promoter/enhancer is used (U.S. Patent 5,168,062). The promoter may be cell-specific and permit substantial transcription of the polynucleotide only in predetermined cells.

- 25 The polynucleotide is introduced directly into the host either as "naked" DNA (U.S. Patent 5,580,859) or formulated in compositions with agents which facilitate immunization, such as bupivacaine and other local anesthetics (U.S. Patent 5,593,972) and cationic polyamines (U.S. Patent 6,127,170).

- In this polynucleotide immunization procedure, the polypeptides of the invention are expressed on a transient basis *in vivo*; no genetic material is inserted or
- 30 integrated into the chromosomes of the host. This procedure is to be distinguished from gene therapy, where the goal is to insert or integrate the genetic material of interest into the chromosome. An assay is used to confirm that the polynucleotides

administered by immunization do not give rise to a transformed phenotype in the host (U.S. Patent 6,168,918).

#### H. USES AND METHODS OF THE INVENTION

5       The *Streptococcus pneumoniae* polynucleotides, polypeptides, polypeptide homologues, modulators, adjuvants, and antibodies described in this invention can be used in methods of treatment, diagnostic assays particularly in disease identification, drug screening assays and monitoring of effects during clinical trials. The isolated polynucleotides of the invention can be used to express *Streptococcus*  
10 *pneumoniae* polypeptides (e.g., via a recombinant expression vector in a host cell or in polynucleotide immunization applications) and to detect *Streptococcus pneumoniae* mRNA (e.g., in a biological sample). Moreover, the anti-*Streptococcus pneumoniae* antibodies of the invention can be used to detect and isolate a *Streptococcus pneumoniae* polypeptide, particularly fragments of a *Streptococcus*  
15 *pneumoniae* polypeptides present in a biological sample, and to modulate *Streptococcus pneumoniae* polypeptide activity.

The invention provides immunogenic compositions comprising polypeptides having an amino acid sequence chosen from one of SEQ ID NO:216 through SEQ ID NO:430 or SEQ ID NO: 592 through SEQ ID NO: 752, a biological equivalent thereof  
20 or a fragment thereof. The immunogenic composition may further comprise a pharmaceutically acceptable carrier, as outlined in section G. In certain preferred embodiments, the immunogenic composition will comprise one or more adjuvants.

In another embodiment, the invention provides immunogenic compositions comprising a polynucleotide having a nucleotide sequence chosen from one of SEQ  
25 ID NO:1 through SEQ ID NO:215 or SEQ ID NO: 431 through SEQ ID NO: 591, wherein the polynucleotide is comprised in a recombinant expression vector. Preferably the vector is plasmid DNA. Of course, the polynucleotide may further comprise heterologous nucleotides, e.g., the polynucleotide is operatively linked to one or more gene expression regulatory elements, and further comprise one or more  
30 adjuvants. In a preferred embodiment, the immunogenic polynucleotide composition directs the expression of a neutralizing epitope of *Streptococcus pneumoniae*.

Provided also are methods for immunizing a host against *Streptococcus pneumoniae* infection. In a preferred embodiment, the host is human. Thus, a host

or subject is administered an immunizing amount of an immunogenic composition comprising a polypeptide having an amino acid sequence chosen from one of SEQ ID NO:216 through SEQ ID NO:430 or SEQ ID NO: 592 through 752, a biological equivalent thereof or a fragment thereof and a pharmaceutically acceptable carrier.

- 5 An immunizing amount of an immunogenic composition can be determined by doing a dose response study in which subjects are immunized with gradually increasing amounts of the immunogenic composition and the immune response analyzed to determine the optimal dosage. Starting points for the study can be inferred from immunization data in animal models. The dosage amount can vary depending upon  
10 specific conditions of the individual. The amount can be determined in routine trials by means known to those skilled in the art.

- An immunologically effective amount of the immunogenic composition in an appropriate number of doses is administered to the subject to elicit an immune response. Immunologically effective amount, as used herein, means the  
15 administration of that amount to a mammalian host (preferably human), either in a single dose or as part of a series of doses, sufficient to at least cause the immune system of the individual treated to generate a response that reduces the clinical impact of the bacterial infection. Protection may be conferred by a single dose of the immunogenic composition or vaccine, or may require the administration of several  
20 doses, in addition to booster doses at later times to maintain protection. This may range from a minimal decrease in bacterial burden to prevention of the infection. Ideally, the treated individual will not exhibit the more serious clinical manifestations of the *Streptococcus pneumoniae* infection. The dosage amount can vary depending upon specific conditions of the individual, such as age and weight. This amount can  
25 be determined in routine trials by means known to those skilled in the art.

#### I. DIAGNOSTIC ASSAYS

- The invention provides methods for detecting the presence of a  
*Streptococcus pneumoniae* polypeptide or *Streptococcus pneumoniae*  
30 polynucleotide, or fragment thereof, in a biological sample. The method involves contacting the biological sample with a compound or an agent capable of detecting a *Streptococcus pneumoniae* polypeptide or mRNA such that the presence of the *Streptococcus pneumoniae* polypeptide/encoding nucleic acid molecule is detected

in the biological sample. A preferred agent for detecting *Streptococcus pneumoniae* mRNA or DNA is a labeled or labelable oligonucleotide probe capable of hybridizing to *Streptococcus pneumoniae* mRNA or DNA. The nucleic acid probe can be, for example, a full-length *Streptococcus pneumoniae* polynucleotide of one of SEQ ID NO: 1 through SEQ ID NO:215 or SEQ ID NO: 431 through SEQ ID NO: 591, a complement thereof, or a fragment thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to *Streptococcus pneumoniae* mRNA or DNA. Alternatively, the sample can be contacted with an oligonucleotide primer of a *Streptococcus pneumoniae* polynucleotide of one of SEQ ID NO: 1 through SEQ ID NO:215 or SEQ ID NO: 431 through SEQ ID NO: 591, a complement thereof, or a fragment thereof, in the presence of nucleotides and a polymerase, under conditions permitting primer extension.

A preferred agent for detecting *Streptococcus pneumoniae* polypeptide is a labeled or labelable antibody capable of binding to a *Streptococcus pneumoniae* polypeptide. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or F(ab')<sub>2</sub>) can be used. The term "labeled or labelable," with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labelling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect *Streptococcus pneumoniae* mRNA, DNA, or protein in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of *Streptococcus pneumoniae* mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of *Streptococcus pneumoniae* polypeptide include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. Alternatively, *Streptococcus pneumoniae* polypeptides can be detected *in vivo* in a

subject by introducing into the subject a labeled anti-*Streptococcus pneumoniae* antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

The polynucleotides according to the invention may also be used in analytical

- 5 DNA chips, which allow sequencing, the study of mutations and of the expression of genes, and which are currently of interest given their very small size and their high capacity in terms of number of analyses.

The principle of the operation of these chips is based on molecular probes, most often oligonucleotides, which are attached onto a miniaturized surface, 10 generally of the order of a few square centimeters. During an analysis, a sample containing fragments of a target nucleic acid to be analysed, for example DNA or RNA labelled, for example, after amplification, is deposited onto the DNA chip in which the support has been coated beforehand with probes. Bringing the labelled target sequences into contact with the probes leads to the formation, through 15 hybridization, of a duplex according to the rule of pairing defined by J.D. Watson and F. Crick. After a washing step, analysis of the surface of the chip allows the effective hybridizations to be located by means of the signals emitted by the labels tagging the target. A hybridization fingerprint results from this analysis which, by appropriate computer processing, will make it possible to determine information such as the 20 presence of specific fragments in the sample, the determination of sequences and the presence of mutations.

The chip consists of a multitude of molecular probes, precisely organized or arrayed on a solid support whose surface is miniaturized. It is at the centre of a system where other elements (imaging system, microcomputer) allow the acquisition 25 and interpretation of a hybridization fingerprint.

The hybridization supports are provided in the form of flat or porous surfaces (pierced with wells) composed of various materials. The choice of a support is determined by its physicochemical properties, or more precisely, by the relationship between the latter and the conditions under which the support will be placed during 30 the synthesis or the attachment of the probes or during the use of the chip. It is therefore necessary, before considering the use of a particular support, to consider characteristics such as its stability to pH, its physical strength, its reactivity and its chemical stability as well as its capacity to nonspecifically bind nucleic acids.



Materials such as glass, silicon and polymers are commonly used. Their surface is, in a first step, called "functionalization", made reactive towards the groups which it is desired to attach thereon. After the functionalization, so-called spacer molecules are grafted onto the activated surface. Used as intermediates between the surface and the probe, these molecules of variable size render unimportant the surface properties of the supports, which often prove to be problematic for the synthesis or the attachment of the probes and for the hybridization.

Among the hybridization supports, there may be mentioned glass which is used, for example, in the method of *in situ* synthesis of oligonucleotides by photochemical addressing developed by the company Affymetrix (E.L. Sheldon, 1993), the glass surface being activated by silane. Genosensor Consortium (P. Mérel, 1994) also uses glass slides carrying wells 3 mm apart, this support being activated with epoxysilane.

The probes according to the invention may be synthesized directly *in situ* on the supports of the DNA chips. This *in situ* synthesis may be carried out by photochemical addressing (developed by the company Affymax (Amsterdam, Holland) and exploited industrially by its subsidiary Affymetrix (United States), or based on the VLSIPS (very large scale immobilized polymer synthesis) technology (S.P.A. Fodor *et al.*, 1991), which is based on a method of photochemically directed combinatory synthesis. The principle of which combines solid-phase chemistry, the use of photolabile protecting groups and photolithography.

The probes according to the invention may be attached to the DNA chips in various ways such as electrochemical addressing, automated addressing or the use of probe printers (T. Livache *et al.*, 1994; G. Yershov *et al.*, 1996; J. Derisi *et al.*, 1996, and S. Borman, 1996).

The revealing of the hybridization between the probes of the invention, deposited or synthesized *in situ* on the supports of the DNA chips, and the sample to be analysed, may be determined, for example, by measurement of fluorescent signals, by radioactive counting or by electronic detection.

The use of fluorescent molecules such as fluorescein constitutes the most common method of labelling the samples. It allows direct or indirect revealing of the hybridization and allows the use of various fluorochromes.

Affymetrix currently provides an apparatus or a scanner designed to read its

Gene Chip™ chips. It makes it possible to detect the hybridizations by scanning the surface of the chip in confocal microscopy (R.J. Lipshutz *et al.*, 1995).

The nucleotide sequences according to the invention may also be used in DNA chips to carry out the analysis of the expression of the *Streptococcus pneumoniae* genes. This analysis of the expression of *Streptococcus pneumoniae* genes is based on the use of chips where probes of the invention, chosen for their specificity to characterize a given gene, are present (D.J. Lockhart *et al.*, 1996; D.D. Shoemaker *et al.*, 1996). For the methods of analysis of gene expression using the DNA chips, reference may, for example, be made to the methods described by D.J. Lockhart *et al.* (1996) and Sosnowsky *et al.* (1997) for the synthesis of probes *in situ* or for the addressing and the attachment of previously synthesized probes. The target sequences to be analysed are labelled and in general fragmented into sequences of about 50 to 100 nucleotides before being hybridized onto the chip. After washing as described, for example, by D.J. Lockhart *et al.* (1996) and application of different electric fields (Sosnowsky *et al.*, 1997), the labelled compounds are detected and quantified, the hybridizations being carried out at least in duplicate. Comparative analyses of the signal intensities obtained with respect to the same probe for different samples and/or for different probes with the same sample, determine the differential expression of RNA or copy numbers of DNA derived from the sample.

The nucleotide sequences according to the invention may, in addition, be used in DNA chips where other nucleotide probes specific for other microorganisms are also present, and may allow the carrying out of a serial test allowing rapid identification of the presence of a microorganism in a sample.

Accordingly, the subject of the invention is also the nucleotide sequences according to the invention, characterized in that they are immobilized on a support of a DNA chip.

The DNA chips, characterized in that they contain at least one nucleotide sequence according to the invention, immobilized on the support of the said chip, also form part of the invention.

The chips will preferably contain several probes or nucleotide sequences of the invention of different length and/or corresponding to different genes so as to identify, with greater certainty, the specificity of the target sequences or the desired

mutation in the sample to be analysed.

Accordingly, the analyses carried out by means of primers and/or probes according to the invention, immobilized on supports such as DNA chips, will make it possible, for example, to identify, in samples, mutations linked to variations such as  
5 intraspecies variations. These variations may be correlated or associated with pathologies specific to the variant identified and will make it possible to select the appropriate treatment.

The invention thus comprises a DNA chip according to the invention, characterized in that it contains, in addition, at least one nucleotide sequence of a  
10 microorganism different from *Streptococcus pneumoniae*, immobilized on the support of the said chip; preferably, the different microorganism will be chosen from an associated microorganism, a bacterium of the *Streptococcus* family, and a variant of the species *Streptococcus pneumoniae*.

The principle of the DNA chip as explained above, may also be used to  
15 produce protein "chips" on which the support has been coated with a polypeptide or an antibody according to the invention, or arrays thereof, in place of the DNA. These protein "chips" make it possible, for example, to analyse the biomolecular interactions (BIA) induced by the affinity capture of target analytes onto a support coated, for example, with proteins, by surface plasma resonance (SPR). Reference may be  
20 made, for example, to the techniques for coupling proteins onto a solid support which are described in International Application EP 524 800 or to the methods describing the use of biosensor-type protein chips such as the BIAcore-type technique (Pharmacia) (Arlinghaus *et al.*, 1997, Krone *et al.*, 1997, Chatelier *et al.*, 1995). These polypeptides or antibodies according to the invention, capable of specifically  
25 binding antibodies or polypeptides derived from the sample to be analysed, may thus be used in protein chips for the detection and/or the identification of proteins in samples. The said protein chips may in particular be used for infectious diagnosis and may preferably contain, per chip, several polypeptides and/or antibodies of the invention of different specificity, and/or polypeptides and/or antibodies capable of  
30 recognizing microorganisms different from *Streptococcus pneumoniae*.

Accordingly, the subject of the present invention is also the polypeptides and the antibodies according to the invention, characterized in that they are immobilized on a support, in particular of a protein chip.

The protein chips, characterized in that they contain at least one polypeptide or one antibody according to the invention immobilized on the support of the said chip, also form part of the invention.

- 5 The invention comprises, in addition, a protein chip according to the invention, characterized in that it contains, in addition, at least one polypeptide of a microorganism different from *Streptococcus pneumoniae* or at least one antibody directed against a compound of a microorganism different from *Streptococcus pneumoniae*, immobilized on the support of the chip.

- 10 The invention also relates to a kit or set for the detection and/or the identification of bacteria belonging to the species *Streptococcus pneumoniae* or to an associated microorganism, or for the detection and/or the identification of a microorganism characterized in that it comprises a protein chip according to the invention.

- 15 The present invention also provides a method for the detection and/or the identification of bacteria belonging to the species *Streptococcus pneumoniae* or to an associated microorganism in a biological sample, characterized in that it uses a nucleotide sequence according to the invention.

- 20 The invention also encompasses kits for detecting the presence of a *Streptococcus pneumoniae* polypeptide in a biological sample. For example, the kit comprises reagents such as a labeled or labelable compound or agent capable of detecting *Streptococcus pneumoniae* polypeptide or mRNA in a biological sample; means for determining the amount of *Streptococcus pneumoniae* polypeptide in the sample; and means for comparing the amount of *Streptococcus pneumoniae* polypeptide in the sample with a standard. The compound or agent is packaged in a  
25 suitable container. The kit further comprises instructions for using the kit to detect *Streptococcus pneumoniae* mRNA or protein.

- In certain embodiments, detection involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent 4,683,195 and U.S. Patent 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain  
30 reaction (LCR). This method includes the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a *Streptococcus pneumoniae* polynucleotide under conditions

such that hybridization and amplification of the *Streptococcus pneumoniae*-polynucleotide (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample.

- 5 All patents and publications cited herein are hereby incorporated by reference.

#### EXAMPLES

- 10 The following examples are carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail. The following examples are presented for illustrative purpose, and should not be construed in any way limiting the scope of this invention.

#### EXAMPLE 1

##### 15 BIOINFORMATICS AND GENE MINING OF *STREPTOCOCCUS PNEUMONIAE*

- The genomic sequence of *Streptococcus pneumoniae* was downloaded from The Institute for Genomic Research (TIGR) website and novel open reading frames (ORFs) were determined in the following manner. An ORF was defined as having one of three potential start site codons, ATG, GTG or TTG and one of three potential stop codons, TAA, TAG or TGA. The inventors used a unique set of two ORF finder algorithms: GLIMMER (Salzberg *et al.*, 1998) and inventors' assignee's program to enhance the efficiency for finding "all" ORFs. In order to evaluate the accuracy of the ORFs determined, a program developed by inventors' assignee called DiCTION was employed that uses a discrete mathematical cosine function to assign a score for each ORF. An ORF with a DiCTION score > 1.5 is considered to have a high probability of encoding a protein product. The minimum length of an ORF predicted by the two ORF finding algorithms was set to 225 nucleotides (including stop codon) which would encode a protein of 74 amino acids. As a final search for remnants of ORFs, all noncoding regions > 75 nucleotides were searched against the public protein databases (described below) using tBLASTn. This helped to identify regions of genes that contain frameshifts (Mejlhede *et al.*, 1999) or fragments of genes that might have a role in causing antigenic variation (Fraser *et al.*, 1997). A graphical analysis program developed by inventors' assignee also allowed the inventors to see

all six reading frames and the location of the predicted ORFs relative to the genomic sequence for further inspection. This helped to eliminate those ORFs that have large overlaps with other ORFs, although there are known cases of ORFs being totally embedded within other ORFs (Loessner *et al.*, 1999; Hernandez-Sanchez *et al.*, 5 1998).

The initial annotation of the *Streptococcus pneumoniae* ORFs was performed using the BLAST (v. 2.0) Gapped search algorithm, Blastp, to identify homologous sequences (Altschul *et al.*, 1997). A cutoff 'e' value of anything  $< e^{-10}$  was considered significant. Other search algorithms such as FASTA or PSI-BLAST were used as 10 needed. The non-redundant protein sequence database used for the homology searches consisted of GenBank, SWISS-PROT (Bairoch and Apweiler, 2000), PIR (Barker *et al.*, 2001), and TREMBL (Bairoch and Apweiler, 2000) database sequences updated daily. ORFs with a Blastp result of  $> e^{-10}$  were considered to be unique to *Streptococcus pneumoniae*.

15 A keyword search of the entire BLAST results was carried out using known or suspected target genes for immunogenic compositions, as well as words that identified the location of a protein or function.

Several parameters were used to determine grouping of the predicted proteins. Proteins destined for translocation across the cytoplasmic membrane 20 encode a leader signal (also called signal sequence) composed of a central hydrophobic region flanked at the N-terminus by positively charged residues (Pugsley, 1993). A program, called SignalP, identifies signal peptides and their cleavage sites (Nielsen *et al.*, 1997). To predict protein localization in bacteria, the software PSORT has been used (Nakal and Kanehisa, 1991). This program uses a 25 neural net algorithm to predict localization of proteins to the 'cytoplasm', 'periplasm', and 'cytoplasmic membrane' for Gram-positive bacteria as well as 'outer membrane' for Gram-negative bacteria. Transmembrane (TM) domains of proteins have been analyzed using the software program TopPred II (Cserzo *et al.*, 1997).

The Hidden Markov Model (HMM) Pfam database of multiple alignments of 30 protein domains or conserved protein regions (Sonnhammer *et al.*, 1997) was used to identify *Streptococcus pneumoniae* proteins that may belong to an existing protein family. Keyword searching of this output was used to help identify additional candidate ORFs that may have been missed by the BLAST search criteria. A

computer algorithm, called HMM Lipo, was developed by inventors' assignee to predict lipoproteins using approximately 131 biologically proven bacterial lipoproteins. This training set was generated from experimentally proven prokaryotic lipoproteins. The protein sequence from the start of the protein to the cysteine amino acid plus the  
5 next two additional amino acids was used to generate the HMM. Using approximately 70 known prokaryotic proteins containing the LPXTG cell wall sorting signal, a HMM (Eddy, 1996) was developed to predict cell wall proteins that are anchored to the peptidoglycan layer (Mazmanian *et al.*, 1999; Navarre and Schneewind, 1999). The model used not only the LPXTG sequence but also  
10 included two features of the downstream sequence, first the hydrophobic transmembrane domain and secondly, the positively charged carboxy terminus. There are also a number of proteins that interact, non-covalently, with the peptidoglycan layer and are distinct from the LPXTG protein class described above. These proteins seem to have a consensus sequence at their carboxy terminus  
15 (Koebnik, 1995). The inventors' assignee has also developed and used a HMM of this region to identify any *Streptococcus pneumoniae* that may fall into this class of proteins.

The proteins encoded by *Streptococcus pneumoniae* identified ORFs were also evaluated for other useful characteristics. A tandem repeat finder (Benson,  
20 1999) identified ORFs containing repeated DNA sequences such as those found in MSCRAMMs (Foster and Hook, 1998) and phase variable surface proteins of *Neisseria meningitidis* (Parkhill *et al.*, 2000). Proteins that contain the Arg-Gly-Asp (RGD) attachment motif, together with integrins that serve as their receptor, constitute a major recognition system for cell adhesion. RGD recognition is one  
25 mechanism used by microbes to gain entry into eukaryotic tissues (Stockbauer *et al.*, 1999; Isberg and Tran Van Nhieu, 1994). However, not all RGD containing proteins mediate cell attachment. It has been shown that RGD containing peptides with a proline at the carboxy end (RGDP) are inactive in cell attachment assays (Pierschbacher and Ruoslahti, 1987) and are excluded. The Geanfammer software  
30 was used to cluster proteins into homologous families (Park and Teichmann, 1998). Preliminary analysis of the family classes has provided novel ORFs within a candidate cluster as well as defining potential protein function.

**EXAMPLE 2**  
**CLONING, EXPRESSION AND ANALYSIS**  
**OF PREDICTED ORF PROTEINS**

**5 MATERIALS AND METHODS**

Growth of *Streptococcus pneumoniae*. *Streptococcus pneumoniae* were grown in Todd Hewitt broth (Difco) supplemented with 0.5% yeast extract. Bacteria were incubated at 35°C in 5% CO<sub>2</sub> without shaking. Mid-log phase cultures (OD<sub>660</sub> approx 0.3) were harvested after approximately 4 hours incubation and cells pelleted  
10 by centrifugation (5,000 x g) at 4°C.

Cloning and expression of predicted ORFs. The predicted ORFs were cloned and expressed in *E. coli* Top10 or BLR(DE3). Expression of each ORF was tested in both pBAD/Thio-TOPO (which contains an arabinose inducible promoter) and pCR-T7/NT-TOPO expression systems (Invitrogen, Carlsbad, CA). Gene specific primers  
15 were designed to amplify, by polymerase chains reaction (PCR), each selected ORF from *Streptococcus pneumoniae* CP1200 (Morrison *et al.*, 1983) genomic DNA purified using the Wizard Genomic DNA purification kit (Promega, Madison, WI). The 5' primers were designed to exclude the predicted signal sequence (as predicted by SignalP) and the 3' primer was designed to either include the stop codon (pCR-T7) or  
20 exclude the stop codon (pBAD). ORFs were amplified in a standard polymerase chain reaction (200 µM each dNTP (Invitrogen), 200 µM each 5' and 3' gene specific primer, 1 µL stock of chromosomal DNA, 2.5U *Pfu* Turbo polymerase (Stratagene, LaJolla, CA) and 1x *Pfu* Turbo reaction buffer in a total volume of 50 µL). Overhanging A's were added to the PCR products by incubation for 10 minutes at  
25 72°C with 1U of Taq DNA polymerase (Roche Diagnostics, Indianapolis, IN). PCR products were cloned into the expression vectors and transformed into *E. coli* TOP10 following manufacturer's TOPO-TA cloning protocol (Invitrogen). Positive clones were identified by PCR using one gene specific primer and one vector specific primer to ensure correct orientation.

30 ORFs cloned into pCR-T7 were transformed into *E. coli* BL21(DE3) for protein expression using the T7 promoter and those cloned into pBAD were kept in TOP10. Protein expression was determined by growing overnight cultures of the positive clones in 2 mL HySoy broth (DMV International Nutritional, Fraser, NY)



supplemented with 100 µg/mL ampicillin. These cultures were then diluted 1:100 into fresh media and grown until  $OD_{600} = 1.0$ . Protein expression was induced with either 2% arabinose (pBAD) or 0.1 mM IPTG (pCRT7). Three hours post-induction, the cells were harvested and protein expression determined by Western blot analysis of whole-cell lysates using either anti-express epitope (pCRT7) or anti-thio (pBAD) antibodies. The best expressing clone (pBAD or pCRT7) was used for protein production and purification.

Fourteen of the ORFs that did not express in either pCRT7 or pBAD were cloned into pET27b(+) (Novagen, Madison, WI). The ORFs were again amplified by PCR and cloned using standard molecular biology techniques into the NcoI and XhoI sites of pET27b(+). Clones were again screened by PCR, and plasmids with the correct insert were transformed into BL21(DE3) and expression tested as described for pCR-T7. Protein expression was determined by Western blot analysis using anti-HSV epitope antibody.

Purification of Soluble His-tag ORF Proteins. Protein was expressed from positive clones in 4 x 1L of media as described above. Cells were harvested by centrifugation, resuspended in 100 mL of Ni Buffer A (20mM Tris, pH 7.5, 150 mM NaCl) and lysed by 2 passages through a French pressure cell at 16,000 psi (SLM Instruments, Inc., Rochester, NY).

For soluble proteins, the cell debris was pelleted by centrifugation at ~9,000 x g and the supernatant was loaded onto an iminodiaceticacid sepharose 6B (Sigma Chemical, St. Louis, MO) column charged with  $Ni^{2+}$ . Unbound proteins were washed from the column with Ni buffer A until  $A_{280}$  of eluate reached a baseline. The bound protein was then eluted with Ni buffer A containing 300 mM imidazole (Sigma Chemical). Purity was estimated by SDS-PAGE.

Samples requiring further purification were concentrated and buffer exchanged over a PD-10 column (Amersham-Pharmacia Biotech, Piscataway, NJ) equilibrated with buffer A (20 mM Tris, pH 8.0). The eluate was loaded onto a Q-sepharose High Performance (Amersham-Pharmacia Biotech) column and eluted with a 0–35% Buffer B (20 mM Tris, pH 8.0, 1M NaCl) gradient. Protein-containing fractions were determined by SDS-PAGE. All protein purification was done using an AKTA Explorer (Amersham-Pharmacia Biotech).

Isolation and Solubilization of Insoluble His-tag fusion proteins. Bacterial cell pellets were suspended at a ratio of 5:1 (buffer volume: pellet wet weight) in 10 mM NaPO<sub>4</sub>/150mM NaCl/pH 7.0 with Complete Protease Inhibitor Cocktail containing EDTA (Roche Diagnostics GmbH, Mannheim, Germany). The cells were disrupted using a Microfluidizer (Microfluidics Corp., Newton, MA) and centrifuged at 21,900 x g for 30 minutes at 4°C. The pellet, containing insoluble His-tag proteins, was subjected to a series of detergent extractions followed by a final solubilization step using 6M urea. The pellet was resuspended in 10 mM NaPO<sub>4</sub>/150 mM NaCl/pH 7.0 containing Complete Protease Inhibitor Cocktail and 1.0% Triton X-100 (TX-100) using the same 5:1 ratio described above. The suspension was stirred at 4°C for 30 minutes and centrifuged at 21,900 x g for 20 minutes at 4°C. The supernatant was removed and stored at 4°C for further analysis. The pellet was subjected to a second TX-100 extraction, as described, and the supernatant removed and stored at 4°C for further analysis. The TX-100 pellet was then resuspended in 10 mM NaPO<sub>4</sub>/150 mM NaCl /pH 7.0 containing Complete Protease Inhibitor Cocktail and 1.0% Zwittergent 3-14 (Z3-14) and stirred at 4°C for a minimum of 1 hour. The suspension was centrifuged at 21,900 x g for 20 minutes at 4°C. The supernatant was removed and stored at 4°C for further analysis. The Z3-14 pellet was resuspended in 100 mM Tris-HCl/6M urea/pH 8.0 and stirred a minimum of 4 hours at room temperature. The suspension was centrifuged at 21,900 x g for 20 minutes at 4°C and the supernatant stored at 4°C for further analysis.

Purification of Solubilized His-tag fusion proteins. Isolated extracts containing His-tag fusion proteins were identified as described by SDS-PAGE and/or Western blot analysis. Chromatography was carried out using POROS MC 20 micron metal chelate Ni<sup>2+</sup> media (Perseptive Biosystems, Framingham, MA) prepared according to the manufacturer. Protein extracts were loaded at approximately 5-10 mg of total protein per mL of column media.

For preparations in which the His-tag proteins were soluble in either the cytosolic fraction or detergent extractions by TX-100 or Z3-14, the material was applied directly to a MC 20 column equilibrated with a minimum of 3 column volumes of 10mM NaPO<sub>4</sub>/150 mM NaCl/pH 7.0 for cytosolic proteins, or the same buffer containing either 1.0% TX-100 or 1.0% Z3-14 for proteins isolated in the TX-100 and Z3-14 extractions respectively. For cytosolic material, unbound proteins were

- washed through the column with a minimum of 5 column volumes of equilibration buffer. For TX-100 or Z3-14 containing extracts, unbound proteins were washed through the column with equilibration buffer containing either 0.05% TX-100 or Z3-14, depending on the solubility characteristics of the particular protein. His-tag fusion
- 5 proteins were eluted using a step gradient of 2 column volumes each of 25 mM, 50 mM, 125 mM, and 250 mM imidazole in 10mM  $\text{NaPO}_4$ /150 mM NaCl/pH 7.0 containing either 0.05% TX-100 or 0.05% Z3-14. Fractions containing His-tag protein were identified by SDS-PAGE and pooled. Imidazole was removed by dialysis into an appropriate buffer. Protein concentration was determined by BCA assay (Pierce)
- 10 and, if necessary, preparations were concentrated by either ultrafiltration using Centriprep YM-10 membranes (Millipore, Bedford, MA) or by applying the material to a smaller MC 20 column, under the conditions described, and eluting with 250 mM imidazole followed by dialysis. Protein purity was estimated by SDS-PAGE and scanning densitometry.
- 15 For preparations in which urea was used to denature and solubilize the protein, the material was diluted 3 fold with 100 mM Tris-HCl/0.05% TX-100/pH 7.5 to give a final urea concentration of 2 M. The material was applied to a MC 20 column equilibrated with a minimum of 3 column volumes of 100 mM Tris-HCl/0.05% TX-100/2 M urea/pH 7.5 and unbound proteins were washed through the column with a
- 20 minimum of 5 column volumes of equilibration buffer. His-tag fusion proteins were eluted using a step gradient of 2 column volumes each of 25 mM, 50 mM, 125 mM, and 250 mM imidazole in 100 mM Tris-HCl/0.05% TX-100/2 M urea pH 7.5. Fractions containing His-tag protein were identified by SDS-PAGE and pooled. Imidazole and urea were removed, and the protein refolded by dialysis into an
- 25 appropriate buffer containing 0.05% TX-100. If necessary, preparations were concentrated by either ultrafiltration using Centriprep YM-10 membranes (Millipore, Bedford, MA) or by applying the material to a smaller MC 20 column, under the conditions described, and eluting with 250 mM imidazole followed by dialysis. Protein purity was estimated by SDS-PAGE and scanning densitometry.
- 30 SDS-PAGE & Western Analysis. SDS-PAGE was carried out as described by Laemmli (Laemmli, 1970), using 10-20% (wt/vol) gradient acrylamide gels (Zaxis, Hudson, OH). Proteins were visualized by staining the gels with Simply Blue Safestain (Invitrogen Life Technologies, Carlsbad, CA). The gels were scanned with

a Personal Densitometer SI (Molecular Dynamics Inc., Sunnyvale, CA) and purities were estimated using the Image Quant software (Molecular Dynamics Inc.).

- Transfer of proteins to polyvinylidene difluoride (PVDF) membranes was accomplished with a semidry electroblotter and electroblot buffers (Owl Separation Systems, Portsmouth, NH). The PVDF membrane, containing the transferred protein, was blocked with 5 % non-fat dry milk prepared in PBS (Blotto) for 30 minutes. The membrane was then probed with one of the following primary antibody preparations at the indicated dilution specific for the individual protein expression system: Invitrogen anti-Xpress (1:5000), Invitrogen anti-thioredoxin (1:2000), Novagen anti-HSV epitope (1:5000), Qiagen anti-4X His (1:5000). The membrane was then washed with Blotto followed by Goat anti-mouse alkaline phosphatase conjugate (1:1500) as the secondary antibody (Biosource International, Camarillo, CA). Western blots were developed with 5-bromo-4-chloro-indolylphosphate-nitroblue tetrazolium (BCIP/NBT) phosphatase substrate system (Kirkegaard and Perry Laboratories, Gaithersburg, MD).

Protein quantitation. Protein concentrations were estimated by the bicinchoninic assay (Pierce, Rockford, IL) with bovine serum albumin as the standard.

- Production of anti-ORF sera in mice. Female Swiss Webster mice (Taconic Farms, Germantown, NY) with ages 6 to 8 weeks old were immunized subcutaneously in the neck at weeks 0, 4, and 6 weeks with purified His tag protein. Two separate immunogenic compositions were prepared with each His-tag protein. One immunogenic composition was prepared with the protein formulated with STIMULON™ QS-21 and a second was prepared with the protein formulated with MPL™. Each dose for one group of mice contained 10 µg of purified protein and 20 µg STIMULON™ QS-21, while each dose for the second group of mice contained 10 µg of the same protein and 50 µg MPL™. Serum samples were collected at weeks 0, 4, 6 and 8. Mice were housed in a specific-pathogen free facility and provided water and food ad-libitum.

- Pneumococcal whole-cell ELISAs. *Streptococcus pneumoniae* strains, either type 3 or type 14, were grown in Todd Hewitt broth (Difco) containing 100 µg/ml streptomycin at 35°C without shaking. The bacteria were grown to mid-log phase ( $OD_{550} < 1.0$ ), and heat inactivated for 1 hour at 60°C. Bacteria were pelleted at

10,000 x g and resuspended in PBS to an  $OD_{950} = 0.1$ . Fifty-five  $\mu$ l of this suspension was then added to each well of 96-well Nunc plates and air dried at room temperature. Plates were stored at 4°C until used.

- 5 Wells were blocked with 150  $\mu$ l/well of PBS containing 5% (wt/vol) dry milk (blocking buffer) for 1 hour. Wells were washed 5 times with PBS in a Skantron washer, and mouse sera diluted in blocking buffer (100  $\mu$ l/well) added. Plates were incubated at room temperature for 2 hours and unbound antibodies removed by washing 5 times with PBS in a Skantron washer. Bound antibodies were detected with 100  $\mu$ l/well of peroxidase-labeled goat anti-mouse IgG (1:1,000 dilution of 10 1mg/ml in PBS; KPL) at room temperature for 2 hours. Plates were washed with PBS as above, and developed with 100  $\mu$ l/well ABTS (KPL) for 25 minutes at room temperature. The reactions were stopped with 100  $\mu$ l/well of 1% SDS and the  $OD_{405}$  of each well read on a VERSAmax microplate reader (Molecular Devices Corp., Sunnyvale, Calif.). Endpoint titers of each test serum were calculated as the inverse of the highest mean dilution giving an  $OD_{405} = 0.1$ . 15

- FACS analysis of *Streptococcus pneumoniae*. Strains type 3 and 19F were grown in Todd-Hewitt broth + 0.5% yeast extract from frozen stocks of  $OD_{600} \sim 1.0$  cells. Incubation was at 37°C for 3 to 4 hours without shaking.  $2-3 \times 10^7$  cells, 100  $\mu$ l of  $OD_{600} = 0.5$  for type3, and 50  $\mu$ l for 19F, were pipetted into a 96-well microtiter plate 20 and spun at 4000 rpm in an Eppendorf tabletop centrifuge for 5 minutes. Supernatant was aspirated and cells were resuspended in 95  $\mu$ l PBS-0.5%BSA-0.1% gelatin. Five  $\mu$ l primary antibody was added, mixed and left incubating on ice for 1 hour. Cells were pelleted as before, washed twice with 100  $\mu$ l buffer and resuspended in 99  $\mu$ l buffer. One  $\mu$ l goat anti-mouse secondary antibody conjugated 25 to Alexa Fluor 488 (Molecular Probes, Eugene, OR) was added to the samples, mixed and left incubating on ice for 30 minutes. Cells were washed as before and resuspended in 100  $\mu$ l buffer. Before analyzing on the FACS Vantage SE unit, samples were diluted to 1 ml with buffer. Samples were read on a Becton Dickinson FACS Vantage unit with an Enterprise II laser. Excitation was at 488nm and emission 30 was detected with a photomultiplier tube using a 530/30 filter. Week 0 antisera were run as background control for the week 8 antisera.

Comparison of message from cells grown *in vitro* and *in vivo*. Messenger RNA (mRNA) levels for specific transcripts can be examined by creating a double

- stranded cDNA from the mRNA using reverse transcriptase. This cDNA is then amplified using standard PCR conditions. The resulting amplification products are thus indicative of the message produced. This technique is useful for comparing the expression of specific transcripts under varying environmental conditions, such as growth in culture flasks versus growth *in vivo*.

- Preparation of RNA from cells grown *in vitro*. *In vitro* grown *Streptococcus pneumoniae* serotypes were grown to log phase in 60 ml THB -0.5%YE at 37°C with 5% CO<sub>2</sub>. Bacterial cells were harvested by centrifugation at 1000 x g for 15 minutes at 4°C. The supernatant was aspirated and the cells were resuspended in 1ml
- 10 RNAlater (Ambion, Austin, TX) and stored for >1 hour at 4°C. The cells were then centrifuged in a microfuge for 5 minutes at 8000 x g. The supernatant was aspirated and the cells were resuspended in 100 µl 10% deoxycholate (DOC). 1100 µl of RNAZOL B (Tel-Test, Inc) were then added and the suspension mixed briefly by inversion. 120 µl of CHCl<sub>3</sub> were then added, the sample mixed by inversion and then
- 15 centrifuged in a microfuge at full speed for 10 minutes at 4°C. The aqueous layer was removed and the RNA was precipitated by addition of an equal volume of 2-propanol. The RNA was incubated at 4°C for >1 hour and then centrifuged in a microfuge at full speed for 10 minutes at room temperature. The supernatant was aspirated and the RNA was washed with 75% ETOH and recentrifuged for 5 minutes.
- 20 The supernatant was aspirated and the RNA was resuspended in 50-100 µl nuclease-free water. DNA was removed from the RNA by treating the sample with RNase-free DNAase (DNA FREE, Ambion) for 20 minutes at 37 °C, followed by inactivation of the enzyme by addition of the DNA FREE chelator. The purity and yield of the RNA was assessed by measuring the absorbance at 260 nm and 280
- 25 nm. Absorbance ratios were typically 1.9-2.0. RNA was stored at -70°C.

- Preparation of RNA from cells grown *in vivo*. *In vivo* grown *Streptococcus pneumoniae* serotypes were harvested from sealed dialysis tubing incubated in the peritoneal cavities of Sprague-Dawley rats as described by Orihuela *et al.* (2000). Log phase *Streptococcus pneumoniae* cells were prepared as described above and
- 30 resuspended to 10<sup>6</sup> cfu/ml in RPMI media (Celltech) supplemented with 0.4% glucose. One ml of the cell suspension was sealed in a PVDF dialysis membrane with a 80,000 M<sub>w</sub> cutoff (SpectraPor). Two such bags were implanted intraperitoneally in 400g Sprague Dawley rats (Taconic). The bags remained in the

rats for 22 hours, after which the rats were terminated and the bags were harvested. RNA was prepared from the intraperitoneally grown cells as described above.

- RT-PCR to examine message levels. Specific message for each candidate gene was amplified out from RNA prepared from both *in vitro* and *in vivo* grown cells using RT-PCR. For each reaction, 0.5 µg RNA was incubated with 0.25 µM of the reverse mining primer for 3 minutes at 75°C, then cooled on ice and transferred to 44°C. The message was reverse transcribed using the RETROscript (Ambion) kit according to the manufacturer's directions. ReddyMix (ABgene) was used according to the manufacturer's directions to amplify each message from 2-5 µl of the sample, using 0.25 µM of the above reverse primer and the forward mining primer. Following amplification, 10 µl of the amplified product was electrophoresed on a 1% agarose gel.

## RESULTS

- Cloning of ORFs into expression vectors.** Fifty-nine ORFs were selected for cloning and expression based on prediction of surface exposure from genomic analysis as described above. These ORFs were amplified by PCR and cloned into the expression vectors as described in Materials and Methods. The ORFs were cloned into pBAD/Thio-TOPO and pCR-T7/NT-TOPO. Both vectors fuse a hexahistidine tag and a unique epitope to facilitate purification and identification by western blot respectively. The pBAD vector also fuses a thioredoxin moiety to the cloned protein to enhance solubility.

- Expression of ORFs in *E. coli*.** The genes encoding all 59 ORFs were induced in the appropriate host *E. coli* strains and examined for expression by SDS-PAGE and western blot analysis of whole cell extracts. Of the 59 ORFs, a total of 24 (41%) were expressed at detectable levels. Fourteen of the ORFs that did not express in either of the expression vectors were cloned into pET27b(+) which fuses a hexahistidine tag to the C-terminus and a PelB leader sequence at the N-terminus of the protein. One of the 14 ORFs cloned into pET27b(+) expressed protein.

- Purification of Expressed ORF Proteins.** All of the expressed ORFs contained a 6X His motif to aid in purification. Initial purification of all of the proteins was done using a Ni containing resin according to manufacturer's directions. Twenty of the expressed ORF proteins were purified to acceptable levels of homogeneity for

immunization studies using this affinity purification (Table 17). Specific purification conditions used are detailed in Materials and Methods and in Table 17. Thirteen of the 20 ORF proteins were used to immunize mice and obtain antisera specific for the expressed protein.

5

**Table 17**  
**Purification of Expressed *S. pneumoniae* ORF Proteins**

ORF #	[Protein] mg/ml	Total Protein mg	Purity %	Final Buffer	"PSORT" PREDICTED Location	Location in <i>E. coli</i>
75	0.52	6.8	94%	PBS/1mM EDTA pH 7.4	Outer membrane	Cytosol
2615	0.42	16.8	80%	PBS/1mM EDTA pH 7.4	Outer membrane	Cytosol
3039	0.53 (0.14)	2.91	82%	0.1M Tris/150mM NaCl/ 0.05%Zw3-14/1mM EDTA pH 8.0	Outer membrane	Inclusion Bodies
1143	1.4	196	92%	PBS/0.05%tx-100/ 1mM EDTA pH 7.4	Inner membrane	Inclusion Bodies
1835	0.5 (0.2)	10.5	91.3%	PBS/0.05%tx-100/ 1mM EDTA pH 7.4	Inner membrane	Inclusion Bodies
1568	1.0	5.0	>85%	PBS/0.05%tx-100/ 1mM EDTA pH 7.4	Inner membrane	Inclusion Bodies
2271	4.9	122.5	>90%	PBS, pH 7.4	Inner Membrane	Cytosol
2621	1.5	4.5	>90%	PBS, pH 7.4	Inner Membrane	Cytosol
1104	2.0	-	85-90%	PBS, pH 7.4	Outer Membrane	Cytosol
935	0.1	.5	85%	50mM Glycine-NaOH/ 150mM NaCl/ 0.05%Z3-14 pH 10.0	Outer membrane	Inclusion Bodies



3361	1.67	3.34	98%	PBS/1mM EDTA pH 7.4	Inner membrane	Cytosol
339	0.91 (0.91)	127.4 (27.3)	93.2% (80.8% )	PBS/0.05%tx-100/ 1mM EDTA pH 7.4	Inner Membrane	Inclusion Bodies
2322	0.55 (0.23)	2.5 (0.92)	90%	BS/0.05%tx-100/ 1mM EDTA pH 7.4	Inner Membrane	Inclusion Bodies
1476	1.2 (0.6)	9.6	>80%	PBS/0.05%tx-100/ 1mM EDTA pH 7.4	Inner Membrane	Inclusion Bodies
3115	0.2 (0.5)	2.8	>85%	PBS/0.05%tx-100/ 1mM EDTA pH 7.4	Inner Membrane	Inclusion Bodies
132	4.6	460	95%	PBS pH 7.4	-	Cytosol
3386	3.1	27	85%	PBS pH 7.4	Inner Membrane	Cytosol
2112	0.6	1.8	85%	PBS pH 7.4	Inner Membrane	Cytosol
916	0.26	1.3	>85%	PBS 0.05% Tx-100 pH 7.4	-	Inclusion Bodies
3373	0.97	1.9	84%	PBS 0.05% Z3-14 pH 7.4	Inner Membrane	Inclusion Bodies

**Expression of ORF proteins in *Streptococcus pneumoniae* whole cell lysates.** To determine if the ORFs are being expressed in *Streptococcus pneumoniae*, whole cell lysates of *in vitro* grown cells were probed with the antisera in Western blot analysis. Each antiserum was reactive with the purified recombinant protein as a positive control (data not shown). Whole cell lysates from *Streptococcus pneumoniae* strains type 3, type 14, and type 19F were examined in Western blot, and the results are summarized in Table 18. Proteins from three of the ORFs were undetectable or barely detectable in all of the strains tested. Proteins from eight of the ORFs were expressed in at least 2 of the strains, while proteins from two ORFs were detected in only one of the three strains examined. These results demonstrate that the majority of the proteins from these ORFs were expressed in late log, early stationary phase *Streptococcus pneumoniae*, and that some strains may not express detectable amounts of each ORF at the time point examined.

**Table 18**  
**Whole Cell ELISA and Western Blot Expression Data for *S. pneumoniae* ORFs**

		Whole Cell ELISA		Western Blot Expression <i>In vitro</i>			FACS Analysis	
Vaccine (10 µg)	Adjuvant (20 µg)	Type 3	Type 14	Type 3	Type 14	Type 19F	Type 3	Type 19F
2615	QS21	<200	<200	-	-	-	-	-
3039	QS21	<200	<200	+	++	++	-	-
75	QS21	256	<200	+++	+++	+++	+	-
1568	QS21	4,018	<200	++	+++	+++	-	-
1143	QS21	779	<200	+	++	+	+	-
1835	QS21	202	<200	-	+/-	-	+	-
2271	QS21	442	<200	+++	+++	+++	+	-
2621	QS21	739	<200	++	+	-	++	-
1104	QS21	409	<200	+++	+++	+++	+	-
339	QS21	<200	<200	-	+/-	-	-	ND
2322	QS21	<200	<200	-	-	+/-	-	ND
3361	QS21	<200	<200	-	+	+	+	ND
935	QS21	<200	<200	-	-	-	-	ND
Standard		~45,000	~10,000	ND	ND	ND		

- 5 **Surface exposure of ORF proteins: Whole Cell ELISA.** The 13 antisera against the recombinant ORF proteins were tested for surface reactivity by whole cell ELISA against two strains of *Streptococcus pneumoniae*, type 3 and type 14. The results are shown in Table 18. Seven of the 13 antisera gave detectable whole cell titers against type 3 *Streptococcus pneumoniae*, while none of them gave detectable
- 10 titers against the type 14 strain. When anticapsular serum was tested against the homologous capsular serotype, the titer against the type 14 strain was much lower than that against the type 3 strain (see row labeled "standard" in Table 18). This result indicated that there might have been sensitivity issues with the type 14 whole cell ELISA, because the Western blot data clearly demonstrate that type 14
- 15 *Streptococcus pneumoniae* do express the majority of the proteins of the ORFs (Table 18). The whole cell ELISA titers of antiserum against the proteins of ORF 75 (SEQ ID NO:218), ORF 1104 (SEQ ID NO:282), ORF 2621 (SEQ ID NO:363), ORF 1568 (SEQ ID NO:306), ORF 1143 (SEQ ID NO:285), ORF 2271 (SEQ ID NO:343), and ORF 1835 (SEQ ID NO:315) ranged from slightly above background to 20 times

above background. These results indicate that these antisera detect at least some surface exposed epitopes for these ORFs.

**Surface exposure of ORF proteins: FACS Analysis.** The polyclonal antisera against the proteins from ORFs 2615, 3039, 75, 1568, 1143, 1835, 2271, 2621, 1104, 339, 2322, 3361 and 935, were analyzed for surface reactivity with whole *Streptococcus pneumoniae* cells by FACS analysis as described above. The results of the analyses are shown in Table 18. *Streptococcus pneumoniae* type 3 cells showed a 9-fold increase in geometric mean fluorescence intensity when labeled with antiserum to ORF 2621 (SEQ ID NO:363). A less intense fluorescence intensity was detected with antisera directed against the proteins of ORF 1835 (SEQ ID NO:315), ORF 2271 (SEQ ID NO:343), ORF 75 (SEQ ID NO:218), ORF 1143 (SEQ ID NO:285), and ORF 1104 (SEQ ID NO:282). Nine of the antisera tested did not show any detectable surface reactivity with the *Streptococcus pneumoniae* type 19F strain. This may be due to the level of sensitivity of the technique or the capsule of 19F covering the surface exposed proteins more completely under the conditions tested.

**Analysis of ORF mRNA expression *in vitro* vs. *in vivo*.** Forward and reverse mining primers were used to amplify the full length message for several ORFs, identified by mining algorithms as potential vaccine antigens (Example 1), from type 3 and type 14 cells grown under *in vitro* and *in vivo* conditions. In three of the four ORFs examined, message was detected in both *in vitro* and *in vivo* grown cells. For ORFs 1104 (SEQ ID NO:282) and 1568 (SEQ ID NO:306), the detection of message correlated with the presence of an immunoreactive band on a Western blot of whole cell lysates for the same serotypes. However for ORF 2322 (SEQ ID NO:345), message was detected in both serotype 3 and 14, but no immunoreactive band was present for those serotypes, indicating that either the protein was secreted or that the antibodies generated by the recombinant protein did not recognize the native protein. No message was detected for ORF 935 (SEQ ID NO:265) in either growth condition, which correlates with the absence of an immunoreactive band on a Western blot. In a separate experiment, message of the expected size was detected from RNA made from serotype 14 grown *in vitro* for ORFs 1143 (SEQ ID NO:285), 1475 (SEQ ID NO:298), 3039 (SEQ ID NO:380), 2271 (SEQ ID NO:343), 3115 (SEQ ID NO:388) and 3361 (SEQ ID NO:402)(data not shown).

## Discussion

Prediction of surface exposure is a critical step for genomic mining efforts for identifying candidate antigens. The algorithms utilized herein have been shown in the past to have predictive value for selecting candidate ORFs to examine. The results shown here demonstrate the utility of the algorithms for *Streptococcus pneumoniae* and that they represent an advance over the previously utilized algorithms. Here, 7 out of 13 proteins from ORFs tested are shown to be surface exposed by at least two of the techniques employed. These techniques, including whole cell ELISA and FACS analysis of whole *Streptococcus pneumoniae* cells, have different strengths for detection of surface exposed epitopes of proteins. Whole cell ELISA utilizes fixed cells bound to a solid phase support, while FACS analysis uses living *Streptococcus pneumoniae* in liquid suspension. However, the whole cell ELISA is more sensitive than the FACS analysis, and can thus give a more quantitative determination of surface exposed epitopes at low levels of antibody binding. It is not known why the protein of ORF 2621 was so strongly positive in the FACS analysis, yet had a comparatively low whole cell ELISA titer (Table 18). This may be the result of differing growth conditions or the differing detection conditions employed in each of the assays. However, the data are consistent in that the proteins from 6 ORFs that are noted to have surface exposed epitopes all are positive in both assays employed.

The lack of detection of surface exposure in the 19F strain by FACS is puzzling. None of the ORFs had detectable epitopes on the surface of the 19F strain in the FACS technique used, but the majority of them were well expressed in whole cell lysates from this strain (Table 18). This may be due to the unique capsular material of 19F covering the surface exposed proteins, or that the FACS technique is less sensitive against type 19F cells. It is also possible that none of the proteins tested have surface exposed epitopes in type 19F, but this is extremely unlikely, since even antiserum against another known candidate (PhpA protein) (Zhang *et al.*, 2001) that is surface exposed produced much less detectable surface antibody binding in FACS analysis as compared to type 3 cells (data not shown).

The failure to detect surface reactive antibody in the type 14 whole cell ELISA (Table 18) was also most likely due to the growth of the cells or the assay conditions,

because the standard sera employed gave a much lower titer than normally observed.

5       The RT-PCR data serve to reinforce the potential of the candidate proteins from these ORF's. The data show that *Streptococcus pneumoniae* grown either *in vitro* or *in vivo* produce mRNA specific for the ORFs examined. Since it is known that the ORFs are expressed *in vitro*, it is likely that they are also expressed *in vivo* as well. Experiments are in progress to confirm this using whole cell lysates from *in vivo* grown cells.

10       Not every ORF analyzed could be shown to be expressed in *Streptococcus pneumoniae*. For example, a protein from ORF 935 was not detected by Western blot analysis, whole cell ELISA (Table 18), or RT-PCR (data not shown). It may be that ORF 935 is only expressed under "real" *in vivo* conditions or that the sequencing of the region is incorrect and the expressed protein is out of frame with the true protein produced by *Streptococcus pneumoniae*.

15

## EXAMPLE 3

## STREPTOCOCCUS PNEUMONIAE PROTEOME ANALYSIS

## MATERIALS AND METHODS

- 5        Bacteria and media. *S. pneumoniae* type III (ATCC #6303) was obtained from the American Type Culture Collection, Manassas, VA. *S. pneumoniae* type 19F was obtained from Dr. Gerald Schiffman, State University of New York, Brooklyn, NY. A glycerol stock plate on Tryptic Soy Agar II (TSA II)/5.0% sheep blood plate (Becton Dickinson Microbiology Systems, Cockeysville, MD) was prepared and incubated overnight, at 37°C in the presence of 5.0% CO<sub>2</sub>. Cells from each plate were transferred to 20 ml of Todd-Hewitt Broth/0.5% Yeast Extract (THY) and incubated overnight at 37°C with gentle shaking (10 rpm) in the presence of 5.0% CO<sub>2</sub>. For type 3, the culture was then diluted 10 fold with 100 ml of THY. For type 19F, the culture was then diluted 40 fold with 200 ml of THY. Both of these diluted cultures were subsequently incubated under the above conditions. Type 19F required 9 h incubation time to reach a concentration of  $1 \times 10^9$  cells/ml. Type 3 was incubated overnight and its concentration was not determined.

- 10        Isolation of membrane fraction. The bacterial cultures were spun down and washed with PBS/MgSO<sub>4</sub> (30 mM sodium phosphate/150 mM NaCl/1 mM MgSO<sub>4</sub>, pH 6.8). The pellets were resuspended in 4 ml of PBS/MgSO<sub>4</sub> containing 5 µg Lysozyme (Sigma Chemical Co., St. Louis, MO), and 400 µg Mutanolysin (Sigma). The samples were incubated at 37°C for 1 hour with shaking. After the incubation, ~300 units of RNAse Cocktail™ (Ambion Inc., Austin, TX) was added to each sample. The samples were centrifuged at low speed using a tabletop centrifuge (2.5 k rpm, 10 min, at 4°C). The supernatant was subsequently spun at high speed to pellet the membrane fractions using a Beckman (Beckman Instruments, Inc., Palo Alto, CA) Model L8-70M Preparative Ultracentrifuge (60Ti rotor, at 40k rpm, 4°C, 1 h). The supernatant was removed and the membrane pellet was washed with PBS/MgSO<sub>4</sub>.

- 20        Trypsin digestion of excised SDS-PAGE gel bands. Mini SDS-PAGE gels (10 cm x 10 cm) were run with precast 10-20% (w/v, acrylamide) gradient gels (Zaxis, Hudson, OH) at 200 V. The See Blue molecular weight standard used was obtained from Invitrogen, Carlsbad, CA. The gels were stained with Simply Blue Safestain, a colloidal Coomassie Blue G250 stain (Invitrogen) as per manufacturer's instructions.

Each sample lane, in its entirety, was cut into 15 different bands. For each sample, bands representing identical molecular weight areas of the gel from three sample lanes, run next to each other, were collected together for further processing. The gel slices were washed twice with 0.5 ml of 50% (v/v) aqueous HPLC grade acetonitrile (Burdick & Jackson, Muskegon, MI) for 5 min with gentle shaking and stored frozen at -20°C following removal of the wash liquid. Frozen gel bands were thawed and cut into 1 mm cubes and subjected to in-gel trypsin digestion using a DigestPro robot (ABIMED Analysen-Technik GmbH, Langenfeld, Germany). In the configuration used, up to 30 samples could be processed simultaneously. The automated protocol consisted of the following steps in order: reduction of the protein in the gel bands with dithiothreitol, alkylation with iodoacetamide, digestion with trypsin and elution of the peptides. Sequencing Grade Modified Trypsin obtained from Promega Corporation, Madison, WI was used. This trypsin is highly specific for hydrolysis of peptide bonds at the carboxylic sides of lysine and arginine residues. It is modified by reductive methylation to make it extremely resistant to autolysis, which can generate pseudotrypsin with chymotrypsin-like specificity. Specificity is further improved by treatment with L-1-chloro-3-tosylamido-4-phenylbutan-2-one (TPCK) followed by affinity purification. The peptide digests were collected, dried using a SpeedVac (Thermo Savant, Holbrook, NY) to ~10 µl, and subsequently diluted to 50 µl with 0.1 M acetic acid. Samples were transferred to plastic autosampler vials, sealed, and injected using a 5 µl sample loop.

Microcapillary LC-Mass Spectrometry. Mass spectral data were acquired on a Thermo Finnigan LCQ DECA quadrupole ion trap mass spectrometer (Thermo Finnigan, San Jose, CA) equipped with a microcapillary reversed-phase HPLC/micro-electrospray interface. Peptide extracts were analyzed on an automated microelectrospray reversed phase HPLC. The microelectrospray interface consisted of a Pico frit fused silica spray needle, 10 cm length by 75 µm ID, 15 µm orifice diameter (New Objective, Cambridge, Massachusetts) packed with 10 µm C<sub>18</sub> reversed-phase beads (YMC, Wilmington, North Carolina) to a length of 10 cm. The Pico frit needle was mounted in a fiber optic holder (Melles Griot, Irvine, California) held on a base positioned at the front of the mass spectrometer detector. The rear of the column was plumbed through a titanium union to supply an electrical connection for the electrospray interface. The union was connected with a length of

fused silica capillary (FSC) tubing to a FAMOS autosampler (LC-Packings, San Francisco, California) that was connected to an HPLC solvent pump (ABI 140C, Perkin-Elmer, Norwalk, Connecticut). The HPLC solvent pump delivered a flow of 50  $\mu\text{L}/\text{min}$ . which was reduced to 250  $\text{nL}/\text{min}$ . using a PEEK microtight splitting tee (Upchurch Scientific, Oak Harbor, Washington), and then delivered to the autosampler using an FSC transfer line. The HPLC pump and autosampler were each controlled using their internal user programs.

Five microliters of the tryptic digest was separated using the  $\text{C}_{18}$  microcapillary HPLC column eluting directly into the orifice of the mass spectrometer.

Peptides were separated at a flow rate of 250  $\text{nL}/\text{min}$  using a 50 minute gradient of 4-65% (v/v) acetonitrile in 0.1 M acetic acid. Peptide analyses were conducted on the LCQ-DECA ion trap mass spectrometer operating at a spray voltage of 1.5 kV, and using a heated capillary temperature of 140° C. Data were acquired in automated MS/MS mode using the data acquisition software provided with the instrument. As the peptides elute from the HPLC into the mass spectrometer, they are detected and fragmented in a data dependent manner using "dynamic exclusion". In this technique, the ion trap cycles between full scan and collision induced dissociation (CID) mode, first detecting candidate ions, and then collecting them for fragmentation. Decisions about which ions are going to be fragmented are performed by the instrument "on the fly". The ions, once collected, are then added to an exclusion list and are rejected for a window of two minutes. This technique allows the instrument to distribute its time efficiently when presented with analytes of very high complexity. The operation can result in the collection of as many as 1000 to 2000 fragmentation (CID) spectra in a single run. The acquisition method included 1 MS scan (375-600  $\text{m/z}$ ) followed by MS/MS scans of the top 2 most abundant ions in the MS scan. The instrument then conducted a second MS scan (600-1000  $\text{m/z}$ ) followed by MS/MS scans of the top 2 most abundant ions in that scan. The dynamic exclusion and isotope exclusion functions were employed to increase the number of peptide ions that were analyzed (settings: 3  $\text{amu}$  = exclusion width, 3 min = exclusion duration, 30 sec = pre-exclusion duration, 3  $\text{amu}$  = isotope exclusion width). For the current experiment involving 30 samples, the data was collected in a completely automated fashion over 48 hours using the autosampler.



Sequence database search for identification of proteins from CID spectra.

Automated analysis of MS/MS data was performed using the SEQUEST computer algorithm incorporated (Eng, McCormack and Yates, 1994) into the Finnigan Bioworks data analysis package (ThermoFinnigan, San Jose, California) using the protein sequence databases described below. SEQUEST is highly computation intensive, the searches for this study were performed on a dedicated 12 x 600 MHz PC cluster. Peptide matches with Xcorr values greater than 2.0 were loaded into a database for further computational analysis followed by manual verification of the data where necessary (as described below).

## RESULTS AND DISCUSSION

### Proteomics Based Approach

The term 'proteome' has been defined as the proteins expressed by the genome of an organism or tissue. One of the primary goals of analysis of the proteome or proteomics involves identification of proteins in a large-scale high-throughput format. Bacterial membrane preparations constitute a very important source for surface localized proteins, which are likely candidate antigens. A proteomics based approach was taken to identify the protein components of the complex mixture of proteins contained in the membrane fraction of *Streptococcus pneumoniae*. The study of membrane associated proteins offers a very specific and significant challenge for proteomics. The detergents required to keep these proteins in aqueous solution usually interfere with analytical methods. During two-dimensional (2-D) gel electrophoresis, which has been widely used for the analysis of soluble proteins, severe quantitative loss of membrane proteins is often observed. The problem is more severe when immobilized pH gradients are used in the first dimension. To minimize such solubility problems with membrane preparations from some other bacteria, several sample preparations, as well as some novel zwitterionic detergents were tested; all of which were shown to improve the analysis of membrane proteins by 2-D gel electrophoresis. However, applicants believe their success in identifying the major set of outer membrane proteins was quite limited. In view of this, a novel combination of a very simple and a very complex method for identification of the membrane proteome component of *Streptococcus pneumoniae* has been applied, as described below.

In this approach, the membrane preparation was first separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) using a mini gel format, followed by staining of the gel with a colloidal Coomassie blue stain. Fifteen gel bands containing the entire sample lane were excised and the bands digested individually with trypsin. The tryptic peptides were analyzed using microcapillary reversed-phase liquid chromatography-micro-electrospray tandem mass spectrometry (LC-MS/MS) on a Finnigan LCQ Deca quadrupole ion trap mass spectrometer. Tandem mass spectrometry (MS/MS) has been shown to be a powerful approach to analyze proteins (Eng, McCormack and Yates, 1994). In the first step, MS/MS uses a mass analyzer to separate a peptide ion from a mixture of ions, then uses a second step or mass analyzer to activate and dissociate the ion of interest. This process, known as collision-induced dissociation (CID), causes the peptide to fragment at the peptide bonds between the amino acids, and the fragmentation pattern of a peptide is used to determine its amino acid sequence. The SEQUEST computer algorithm (Eng, McCormack and Yates, 1994) was used to search the uninterpreted experimental fragmentation spectra against protein or translated nucleotide sequence databases to identify the proteins present in each gel band. SEQUEST conceptually digests protein sequences in a database into tryptic peptides and then models them into simulated CID spectra using the known rules of peptide fragmentation. SEQUEST then compares these simulated CID spectra against the experimental spectra and returns a list of probable peptide sequences matching the raw data along with different parameters representing the fidelity of the match. For peptides above roughly 800-900 Dalton in size, a single spectrum can uniquely identify a protein.

To obtain sequence information on multiple peptides from the complex mixture generated by trypsin digestion of the SDS-PAGE gel bands, a reversed phase chromatography system was coupled to an electrospray ion trap mass spectrometer. In this system, it is known that high sensitivity (down to sub-femtomole levels) can be attained by minimizing both flow rate and column diameter to concentrate the elution volume and direct as much of the column effluent as possible into the orifice of the mass spectrometer. To maximize the coverage of proteins present in the sample, the data-dependent acquisition feature of the ion trap was employed. Dynamic exclusion was used to prevent reacquisition of tandem mass

spectra of ions once a spectrum had been acquired for a particular m/z value. Use of these data-dependent features dramatically increased the number of peptide ions that were selected for CID analysis.

The LC-MS/MS data acquisition conditions described above typically resulted in fragmentation data for more than 2000 peptide ions for each run. Using the SEQUEST algorithm, this data was correlated against two protein sequence databases. The first one, SnA6F6, contained open reading frames obtained from translation of *Streptococcus pneumoniae* type 4 genome sequence (TIGR4) in all six reading frames with the smallest peptide containing six amino acid residues. The second one, nr, is a non-redundant GenBank protein sequence database. SEQUEST search conditions used trypsin selectivity for both of the searches. The SnA6F6 search allowed a differential search of +16 Dalton for methionine residues to account for peptides displaying oxidation of methionine.

Candidate matches identified by SEQUEST were confirmed using the following procedure. For each peptide, SEQUEST computes a Xcorr value from cross correlation of the experimental MS/MS spectrum with the candidate peptides in the sequence database. The Xcorr is a measure of the similarity of the experimental MS/MS data to that generated from the sequence database. Peptide matches with Xcorr values greater than 2.0 were selected for further analysis and loaded on to an in-house developed system for analysis of SEQUEST data using the commercially available Oracle® relational database system. Since the SEQUEST output is quite complex, applicants incorporated a new scoring algorithm in Oracle® to calculate a match score for each protein identified as follows:

$$\text{Protein Score} = n \sum (\text{Xcorr}/\text{rank})$$

where the rank is that assigned by SEQUEST for each peptide sequence identified from a specific protein sequence in the database and n is the number of unique peptides identified for that protein, since the same peptide may be identified multiple times in an LC-MS/MS experiment. The fragmentation spectra for all moderate or weak assignments by the software used were checked manually by direct examination of the CID spectra for reasonable signal/noise ratio, and the list of matched ions was also examined for reasonable continuity. Generally three or more spectra converging with reasonable Protein Score (usually >25) or Xcorr values (usually >2.5) onto a single database entry constitutes a convincing identification.

The rationale behind the experimental proteomics approach for characterization of membrane associated proteins of *Streptococcus pneumoniae* was that the single SDS-PAGE step circumvented the solubility complications associated with isoelectric focusing in 2-D gel electrophoresis. It also offered a simple fractionation of the membrane preparation according to molecular weight that reduced the complexity of the samples subjected to LC-MS/MS analysis. The combination of these analytical techniques allowed us to separate and obtain sequence information of multiple peptides with high sensitivity over a large concentration range and identify the corresponding proteins by correlation with sequences in databases. As part of this study, a method for the isolation of membrane preparations from *Streptococcus pneumoniae* was also developed. This involved enzymatic digestion of *Streptococcus pneumoniae* cell walls with mutanolysin and lysozyme in a hypotonic buffer followed by differential centrifugation. The twenty-eight ORFs representing surface exposed proteins were also identified by the proteomic approach and are presented in Table 11. The ORFs representing membrane associated proteins and identified by the proteomic approach are presented in Table 12. Table 14 contains all the open reading frames identified from the SnA6F6 database representing the TIGR4 genomic sequence. Table 14 also contains proteins identified from the nr database search which do not originate from the TIGR4 genome.

#### Combination of Genomics and Proteomics Approaches

The ORFs identified by proteomics represent surface localized, surface exposed or membrane associated proteins of *Streptococcus pneumoniae*. Those twenty-eight ORFs that support the putative surface exposed ORFs identified by genomics approaches (i.e., Tables 1-10) are listed in Table 11 and provide further evidence of surface localization of these candidates. The 161 novel ORFs identified by proteomics as membrane associated are listed in Table 12.

#### EXAMPLE 4

##### IMMUNOGOLD LABELING OF *STREPTOCOCCUS PNEUMONIAE* AND LOW VOLTAGE SCANNING ELECTRON MICROSCOPY

Surface exposure of proteins on *Streptococcus pneumoniae* may also be assessed by immunogold labeling of whole bacteria and electron microscopy.

Bacteria cells are labeled as previously described (Olmsted *et al.*, 1993). Briefly, late-log phase bacterial cultures are washed twice, and resuspended to a concentration of  $1 \times 10^8$  cells/ml in 10 mM phosphate buffered saline (PBS) (pH 7.4) and placed on poly-L-lysine coated glass coverslips. Excess bacteria are gently washed from the coverslips and unlabeled samples are placed into fixative (2.0% glutaraldehyde, in a 0.1 M sodium cacodylate buffer containing 7.5% sucrose) for 30 min. Bacteria to be labeled with colloidal gold are washed with PBS containing 0.5% bovine serum albumin, and the pre-immune or hyper-immune mouse polyclonal antibody prepared above applied for 1 hour at room temperature. Bacteria are then gently washed, and a 1:6 dilution of goat anti-mouse conjugated to 18 nm colloidal gold particles (Jackson ImmunoResearch Laboratories, Inc., West Grove, PA) applied for 10 min at room temperature. Finally, all samples are washed gently with PBS, and placed into the fixative described above. The fixative is washed from samples twice for 10 min in 0.1 M sodium cacodylate buffer, and postfixed for 30 min in 0.1 M sodium cacodylate containing 1% osmium tetroxide. The samples are then washed twice with 0.1 M sodium cacodylate, dehydrated with successive concentrations of ethanol, critical point dried by the CO<sub>2</sub> method of Anderson (Anderson, 1951) using a Samdri-780A (Tousimis, Rockville, MD), and coated with a 1-2 nm discontinuous layer of platinum. *Streptococcus pneumoniae* cells are viewed with a LEO 1550 field emission scanning electron microscope operated at low accelerating voltages (1-4.5 keV) using a secondary electron detector for conventional topographical imaging and a high-resolution Robinson backscatter detector to enhance the visualization of colloidal gold by atomic number contrast.

25

#### EXAMPLE 5

##### IN VITRO OPSONOPHAGOCYTOSIS ANALYSIS

An *in vitro* opsonic reaction, that may mimic the *in vivo* reaction, is conducted by incubating together a mixture of *Streptococcus pneumoniae* cells, heat inactivated human serum containing specific antibodies to the pneumococcal strain, and an exogenous complement source. Opsonophagocytosis proceeds during incubation of freshly isolated human polymorphonuclear cells (PMN's) and the antibody/complement/pneumococcal cell mixture. Bacterial cells that are coated with antibody and complement are killed upon opsonophagocytosis. Colony forming units

30

(cfu) of surviving bacteria that escape from opsonophagocytosis are determined by plating the assay mixture. Titers are reported as the reciprocal of the highest dilution that gives  $\geq 50\%$  bacterial killing, as determined by comparison to assay controls. Specimens which demonstrate less than 50% killing at the lowest serum dilution tested (1:8), are reported as having an OPA titer of 4. The highest dilution tested is 1:2560. Samples with  $\geq 50\%$  killing at the highest dilution are repeated, beginning with a higher initial dilution.

The present method is a modification of Gray's method (Gray, B.M. 1990). The assay mixture is assembled in a 96-well microtiter tissue culture plate at room temperature. The assay mixture consists of 10  $\mu\text{L}$  of test serum (a series of two-fold dilutions) heated to  $56^{\circ}\text{C}$  for 30 minutes prior to testing, 10  $\mu\text{L}$  of preclostral bovine serum (complement source) having no opsonic activity for the bacterial test strain, and 20  $\mu\text{L}$  of buffer containing 2000 viable *Streptococcus pneumoniae* organisms. This mixture is incubated at  $37^{\circ}\text{C}$  without  $\text{CO}_2$  for 30 minutes with shaking. Next, 40  $\mu\text{L}$  of human PMNs, freshly prepared from heparinized peripheral blood by dextran sedimentation and Percoll density centrifugation, suspended in buffer at a concentration of  $1 \times 10^6/\text{mL}$  is added. The assay plate(s) are then incubated at  $37^{\circ}\text{C}$  for an additional 90 minutes with vigorous shaking. Aliquots from each well are dispensed onto the upper 1/4 of a 15 x 100 mm blood agar plate. The blood agar plate is tilted while pipetting to allow the liquid suspension to "run" down the plate. Plates are incubated overnight in 5%  $\text{CO}_2$  at  $37^{\circ}\text{C}$ . The viable cfu are counted the following morning. Negative control wells, lacking bacterial cells, test serum, complement and/or phagocytes in appropriate combination are included in each assay. A test serum control, which contains test serum plus bacterial cells and heat inactivated complement, is included for each individual serum. This control can be used to assess whether the presence of antibiotics or other serum components are capable of killing the bacterial strain directly (*i.e.* in the absence of complement or PMN's). A human serum with known opsonic titer is used as a positive human serum control. The opsonic antibody titer for each unknown serum is calculated as the reciprocal of the initial dilution of serum giving 50% cfu reduction compared to the control without serum.

**EXAMPLE 6****INTRANASAL OR PARENTERAL IMMUNIZATION OF****CBA/CAHN MICE PRIOR TO CHALLENGE**

Six-week old, pathogen-free, male CBA/CAHN xid/J (CBA/N) mice are purchased from Jackson Laboratories (Bar Harbor, Maine) and housed in cages under standard temperature, humidity, and lighting conditions. CBA/N mice, at 10 animals per group, are immunized with an appropriate amount of the protein(s) to be tested. For parenteral immunization, the protein is mixed with 100 µg of MPL™ per dose to a final volume of 200 µl in saline and then injected subcutaneously (SC) into mice. All groups receive a booster with the same dose and by the same route 3 and 5 weeks after the primary immunization. Control mice are injected with MPL™ alone. All mice are bled two weeks after the last boosting; sera is then isolated and stored at -20°C. For intranasal (IN) immunization, mice receive three IN immunizations, one week apart. On each occasion, an appropriate dose of the protein to be tested is formulated with 0.1 µg of CT-E29H, a genetically modified cholera toxin that is reduced in enzymatic activity and toxicity (Tebbey *et al.*, 2000), and slowly instilled into the nostril of each mouse in a 10 µl volume. Mice immunized with CT-E29H alone are used as controls. Serum samples are collected one week after the last immunization.

20

**EXAMPLE 7****LD<sub>50</sub> DETERMINATION**

Six or 12-week old CBA/N mice (10 per group) are challenged intranasally (IN) with 10 µl of a suspension of streptomycin resistant type 3 *Streptococcus pneumoniae* diluted to 5 x 10<sup>8</sup> CFU/ml in PBS. Two-fold serial dilutions of this suspension are also tested. The actual doses of bacteria administered are determined by plating dilutions of the inoculum on streptomycin containing tryptic soy agar plates. The LD<sub>50</sub> is calculated by the Reed-Muench method as discussed by Lennette (Lennette, 1995). The LD<sub>50</sub> of 13-week old CBA/N mice with type 3 strain was previously shown to be 1 x 10<sup>5</sup> CFU, while the LD<sub>50</sub> of 6-week old CBA/N mice was 1 x 10<sup>4</sup> CFU.

30

**EXAMPLE 8****CBA/CAHN *XID* MOUSE INTRANASAL CHALLENGE MODEL**

Mice are challenged with either serotype 3 or serotype 14 streptomycin resistant *Streptococcus pneumoniae*. Pneumococci are inoculated into 3 ml of Todd-Hewitt broth containing 100 µg/ml of streptomycin. The culture is grown at 37°C until mid-log phase, then diluted to the desired concentration with Todd-Hewitt broth and stored on ice until use. Each mouse is anesthetized with 1.2 mg of ketamine HCl (Fort Dodge Laboratory, Ft. Dodge, Iowa) by intraperitoneal (IP) injection. The bacterial suspension is inoculated to the nostril of anesthetized mice (10 µl per mouse). The actual dose of bacteria administered is confirmed by plate count. Two or 3 days after challenge, mice are sacrificed, the noses are removed, and homogenized in 3-ml sterile saline with a tissue homogenizer (Ultra-Turax T25, Janke & Kunkel Ika-Labortechnik, Staufen, Germany). The homogenate is 10-fold serially diluted in saline and plated on streptomycin containing TSA plates. Fifty µl of blood collected 2 days post-challenge from each mouse are also plated on the same kind of plates. Plates are incubated overnight at 37°C and then colonies are counted. CBA/N mice are observed daily after challenge, and the mortality is monitored for 14 days.

20

**EXAMPLE 9****INTRANASAL IMMUNIZATION OF BALB/C MICE PRIOR TO CHALLENGE**

Six-week old, pathogen-free, Balb/c mice are purchased from Jackson Laboratories (Bar Harbor, Maine) and housed in cages under standard temperature, humidity, and lighting conditions. BALB/C mice, at 10 animals per group, are immunized with an appropriate amount of the protein to be tested on weeks 0, 2, and 4. On each occasion, the protein being tested is formulated with 0.1 µg of CT-E29H, and slowly instilled into the nostril of each mouse in a 10 µl volume. Mice immunized with Keyhole Limpet Hemocyanin (KLH)-CT-E29H are used as controls. Serum samples are collected 4 days after the last immunization.

30



## EXAMPLE 10

## MOUSE INTRANASAL CHALLENGE MODEL

Balb/c mice are challenged on the sixth day of week 4 (*i.e.*, at approximately 27 days) with  $1 \times 10^5$  CFU's of serotype 3 streptomycin resistant *Streptococcus pneumoniae*. Pneumococci are inoculated into 3 ml of Todd-Hewitt broth containing 100  $\mu$ g/ml of streptomycin. The culture is grown at 37°C until mid-log phase, then diluted to the desired concentration with Todd-Hewitt broth and stored on ice until use. Each mouse is anesthetized with 1.2 mg of ketamine HCl (Fort Dodge Laboratory, Ft. Dodge, Iowa) by i.p. injection. The bacterial suspension is inoculated into the nostril of anesthetized mice (10  $\mu$ l per mouse). The actual dose of bacteria administered is confirmed by plate count. Four days after challenge, mice are sacrificed, the noses removed, and homogenized in 3-ml sterile saline with a tissue homogenizer (Ultra-Turax T25, Janke & Kunkel Ika-Labortechnik, Staufen, Germany). The homogenate is 10-fold serially diluted in saline and plated on streptomycin containing TSA plates. Fifty  $\mu$ l of blood collected 2 days post-challenge from each mouse also is plated on the same kind of plates. Plates are incubated overnight at 37°C and then colonies are counted.

REFERENCES

- International Application No. EP A02323621  
International Application No. EP 0036776  
5 International Application No. EP 0859055  
International Application No. EP 125,023  
International Application No. EP 171,496  
International Application No. EP 171,496  
International Application No. EP 184,187  
10 International Application No. EP 264166  
International Application No. PCT/US86/02269  
U.S. Patent 4,196,265  
U.S. Patent 4,522,811  
U.S. Patent 4,554,101  
15 U.S. Patent 4,683,195  
U.S. Patent 4,683,202  
U.S. Patent 4,736,866  
U.S. Patent 4,816,567  
U.S. Patent 4,870,009  
20 U.S. Patent 4,873,191  
U.S. Patent 4,873,316  
U.S. Patent 4,987,071  
U.S. Patent 5,116,742  
U.S. Patent 5,223,409  
25 U.S. Patent 5,272,057  
U.S. Patent 5,283,317  
U.S. Patent 5,328,470  
U.S. Patent 5,498,531  
U.S. Patent 5,766,844  
30 U.S. Patent 5,789,654  
U.S. Patent 5,798,209  
U.S. Patent 6,201,103  
U.S. SIR No. H1,892

- International Application No. WO 86/01533  
International Application No. WO 90/02809  
International Application No. WO 90/11354  
International Application No. WO 91/01140  
5 International Application No. WO 91/17271  
International Application No. WO 92/01047  
International Application No. WO 92/0968  
International Application No. WO 92/09690  
International Application No. WO 92/15679  
10 International Application No. WO 92/18619  
International Application No. WO 92/20791  
International Application No. WO 93/01288  
International Application No. WO 93/04169  
International Application No. WO94/10300  
15 International Application No. WO 94/16101  
International Application No. WO 97/07668  
International Application No. WO 97/07669  
International Application No. WO 00/63364
- 20 Abravaya *et al.*, *Nucleic Acids Res.*, 23:675-682, 1995.  
Adams *et al.*, *Nature* 355:632-634, 1992.  
Adams *et al.*, *Nature* 377 Supp:3-174, 1995.  
Adams *et al.*, *Science* 252:1651-1656, 1991.  
Altschul *et al.*, "Gapped BLAST and PSI-BLAST: a new generation of protein  
25 database search programs," *Nuc. Acids Res.* 25(17):3389-402, 1997.  
Altschul *et al.*, *J. Molec. Biol.* 215:403-410, 1990.  
Amann *et al.*, *Gene* 69:301-315, 1988.  
Anderson, "Techniques for the preservation of three-dimensional structure in  
preparing specimens for the electron microscope." *Trans. N. Y. Acad. Sci.*  
30 13(130):130-134, 1951.  
Bairoch and Apweiler, *Nucleic Acids Research*, 28:45-48, 2000.  
Baldari *et al.*, *Embo J.* 6:229-234, 1987.  
Banerji *et al.*, *Cell*, 33:729-740; 1983.

- Barker *et al.*, *Nucleic Acids Research*, 29:29-32, 2001.
- Bartel and Szostak, *Science* 261:1411-1418, 1993.
- Bartel *et al.* *Biotechniques* 14:920-924, 1993(b).
- Bartel, "Cellular Interactions and Development: A Practical Approach", pp. 153-179, 1993(a).
- 5 Bateman *et al.*, "The Pfam protein families database," *Nucleic Acid Res.*, 28(1), 263-266, 2000.
- Benson, "Tandem repeats finder: a program to analyze DNA sequences," *Nucleic Acids Res.* 27(2):573-80, 1999.
- 10 Bradley, *Current Opinion in Biotechnology* 2:823-829, 1991.
- Bradley, in "Teratocarcinomas and Embryonic Stem Cells: A Practical Approach," E.J. Robertson, ed., IRL, Oxford, pp. 113-152, 1987.
- Briles *et al.*, "Intranasal immunization of mice with a mixture of the pneumococcal proteins PsaA and PspA is highly protective against nasopharyngeal carriage of *Streptococcus pneumoniae*," *Infect. Immun.* 68(2):796-800, 2000.
- 15 Bunzow *et al.*, *Nature*, 336:783-787, 1988.
- Burge and Karlin, "Prediction of complete gene structures in human genomic DNA." *J. Mol. Biol.* 268:78-94, 1997.
- Butler *et al.*, "Pneumococcal vaccines: history, current status, and future directions," *Am. J. Med.* 107(1A):69S-76S, 1999.
- 20 Byrne and Ruddle, *PNAS* 86:5473- 5477, 1989.
- Calame and Eaton, *Adv. Immunol.* 43:235-275, 1988.
- Campes and Tilghman, *Genes Dev.* 3:537-546, 1989.
- Chen *et al.*, *PNAS* 91:3054-3057, 1994.
- 25 Cohen *et al.*, *Adv. Chromatogr.* 36:127-162, 1996.
- Cotton *et al.*, *PNAS* 85:4397, 1988.
- Cotton, *Mutat. Res.* 285:125-144, 1993.
- Cowan *et al.*, "RGS Proteins: Lessons from the RGS9 subfamily," *Progress in Nucleic Acid Research and Molecular Biology* 65:341-359, 2001.
- 30 Crain *et al.*, "*Streptococcus pneumoniae* coccal surface protein A (PspA) is serologically highly variable and is expressed by all clinically important capsular serotypes of *Streptococcus pneumoniae*," *Infect. Immun.* 58(10):3293-9, 1990.

- Cserzo *et al.*, "Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method," *Protein Engineering* 10(6):673-6, 1997.
- D'Eustachio *et al.*, *Science* 220:919-924, 1983.
- 5 Devereux *et al.*, *Nucleic Acids Research* 12(1):387, 1984.
- Dintilhac, *et al.*, "Competence and virulence of *Streptococcus pneumoniae*: Adc and PsaA mutants exhibit a requirement for Zn and Mn resulting from inactivation of putative ABC metal permeases," *Mol. Microbiol.* 25(4):727-739, 1997.
- Doetschman *et al.*, *J. Embryol. Exp. Morphol.* 87:27-45, 1985.
- 10 Douglas *et al.*, "Antibody response to pneumococcal vaccination in children younger than five years of age," *J. Infect. Dis.* 148:131-137, 1983.
- Eddy, "Hidden Markov models" *Current Opinion in Structural Biology* 6(3):361-5, 1996.
- Edlund *et al.*, *Science* 230:912-916, 1985.
- 15 Eichelbaum, *Clin. Exp. Pharmacol Physiol.* 23(10-11):983-985, 1996.
- Elledge *et al.*, *Proc. Natl. Acad. Sci. USA*, 88:1731-1735, 1991.
- Eng, McCormack and Yates, "An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database," *Journal of the American Society for Mass Spectrometry*, 5:976-989, 1994.
- 20 Fan, Y. *et al.*, *PNAS*, 87:6223-27, 1990.
- Finely *et al.*, *Proc. Natl. Acad. Sci. USA*, 91:12980-12984, 1994.
- Foster and Hook, "Surface protein adhesins of *Staphylococcus aureus*," *Trends Microbiol.* 6(12):484-8, 1998.
- Fraser *et al.*, "Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*" *Nature* 390(6660):580-6, 1997.
- 25 Frohman *et al.*, *Proc. Natl. Acad. Sci. USA* 85, 8998-9002, 1988.
- Gaultier *et al.*, *Nucleic Acids Res.* 15:6625-6641, 1987.
- Gentz *et al.*, *Proc. Natl. Acad. Sci. USA*, 86:821-824, 1989.
- Goldstein and Garau, "30 years of penicillin-resistant *S. pneumoniae*: myth or reality?," *Lancet* 350(9073):233-4.
- 30 Gray, *Conjugate Vaccines Supplement* p694-697, 1990.
- Griffin *et al.*, *Appl. Biochem. Biotechnol.* 38:147-159, 1993.

- Gunnar von Heijne, "Membrane Protein Structure Prediction, Hydrophobicity Analysis and the Positive-inside Rule" *J. Mol. Biol.*, 225:487-494, 1992.
- Harlow and Lane, "Antibodies: A Laboratory Manual," Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1988
- 5 Harper *et al.*, *Cell*, 75:805-816, 1993.
- Haselhoff and Gerlach, *Nature* 334:585-591, 1988.
- Hausdorff *et al.*, "Which pneumococcal serogroups cause the most invasive disease: implications for conjugate vaccine formulation and use, part I," *Clinical Infectious Diseases* 30(1):100-21, 1997.
- 10 Hayashi, *Genet. Anal. Tech. Appl.* 9:73-79, 1992.
- Helene *et al.*, *Ann. N. Y Acad Sci.* 660:27- 36, 1992.
- Helene, *Anticancer Drug Des.* 6(6):569-84, 1991.
- Hepler, "Emerging roles for RGS proteins in cell signalling," *Trends in Pharmacological Sciences* 20:376-382, 1999.
- 15 Hernandez-Sanchez *et al.*, "lambda bar minigene-mediated inhibition of protein synthesis involves accumulation of peptidyl-tRNA and starvation for tRNA," *EMBO Jour.* 17(13):3758-65, 1998.
- Hogan, "Manipulating the Mouse Embryo," *Cold Spring Harbor Laboratory Press*, Cold Spring Harbor, N.Y., 1986.
- 20 Inoue *et al.*, *FEBS Lett.* 215:327-330, 1987(a).
- Inoue *et al.*, *Nucleic Acids Res.* 15:6131-6148, 1987(b).
- Isberg and Tran Van Nhieu, "Binding and internalization of microorganisms by integrin receptors," *Trends in Microbiol.* 2(1):10-4, 1994.
- Iwabuchi *et al.*, *Oncogene* 8:1693-1696, 1993.
- 25 Johnson *et al.*, *Endoc. Rev.*, 10:317-331, 1989.
- Kaufman *et al.*, *EMBO J* 6:187-195, 1987.
- Kessel and Gruss, *Science* 249:3 74-3 79, 1990.
- Klein *et al.*, *Curr. Genet.*, 16:145-152, 1989(b).
- Klein *et al.*, *Curr. Genet.* 13:29-35, 1989(a).
- 30 Koebnik, "Proposal for a peptidoglycan-associating alpha-helical motif in the C-terminal regions of some bacterial cell-surface proteins," *Mol. Microbiol.* 16(6):1269-70, 1995.

- Krapa *et al.*, "Evectins: Vesicular proteins that carry a pleckstrin homology domain and localize to post-Golgi membranes," *Proceedings of the National Academy of Sciences* 96:4633-4368, 1999.
- Kurj an and Herskowitz, *Cell* 933-943, 1982.
- 5 Kyte and Doolittle, *J. Mol. Biol.*, 157:105-132, 1982.
- Lakso *et al.*, *PJVAS* 89:6232-6236, 1992.
- Laemmli, "Cleavage of structural proteins during the assembly of the head of bacteriophage T4," *Nature* (London) 227:680-685, 1970.
- Lefkowitz, *Nature*, 351:353-354, 1991.
- 10 Lennette, "General principles for laboratory diagnosis of viral, rickettsial, and chlamydial infections," p. 17-18, diagnostic procedures for viral, rickettsial, and chlamydial infections, vol. 7th edition, 1995.
- Lewis, "Programmed death in bacteria," *Microbiol. Mol. Biol. Rev.* 64(3):503-14, 2000.
- 15 Li *et al.*, *Cell* 69:915, 1992.
- Linder, *Clin. Chem.* 43(2):254-266, 1997.
- Loessner *et al.*, "Evidence for a holin-like protein gene fully embedded out of frame in the endolysin gene of Staphylococcus aureus bacteriophage 187," *J. Bacteriol.* 181(15):4452-60, 1999.
- 20 Lowry *et al.*, "Protein measurement with the Folin-Phenol reagents," *J. Biol. Chem.* 193:265-275, 1951.
- Lucklow and Summers, *Virology* 170:31-39, 1989.
- Lukashin and Borodovsky, "GeneMark.hmm: new solutions for gene finding," *Nuc. Acids Res.* 26(4):1107-15, 1998.
- 25 Madura *et al.*, *J. Biol. Chem.* 268:12046-1205, 1993
- Maier, *Bioassays* 14(12):807-15, 1992.
- Mansour *et al.*, *Nature* 336:348, 1988
- Maxim and Gilbert, *PNAS* 74:560, 1977.
- Mazmanian *et al.*, "Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall," *Science* 285(5428):760-3, 1999.
- 30 McAtee *et al.*, "Characterization of a Helicobacter pylori vaccine candidate by proteome techniques," *J. Chromatogr. B. Biomed. Sci. Appl.* 714(2):325-33, 1998(c).

- McAtee *et al.*, "Identification of potential diagnostic and vaccine candidates of *Helicobacter pylori* by "proteome" technologies," *Helicobacter* 3(3):163-9, 1998(a).
- 5 McAtee *et al.*, "Identification of potential diagnostic and vaccine candidates of *Helicobacter pylori* by two-dimensional gel electrophoresis, sequence analysis, and serum profiling," *Clin. Diagn. Lab. Immunol* 5(4):537-42, 1998(b).
- McDaniel *et al.*, "Monoclonal antibodies against protease-sensitive pneumococcal antigens can protect mice from fatal infection with *Streptococcus pneumoniae*," *J. Exp. Med.* 160(2):386-97, 1984.
- 10 Mejlhede *et al.*, "Ribosomal -1 frameshifting during decoding of *Bacillus subtilis* cdd occurs at the sequence CGA AAG," *J. Bacteriol.* 181(9):2930-7, 1999.
- Morrison *et al.*, "Isolation and characterization of three new classes of transformation deficient mutants of *Streptococcus pneumoniae* that are defective in DNA transport and genetic recombination," *Journal of Bacteriology*, 156:281-290, 1983.
- 15 Morin *et al.*, *Nucleic Acids Res.*, 21:2157-2163, 1993.
- Myers *et al.*, *Nature* 313:495, 1985(a).
- Myers *et al.*, *Science* 230:1242, 1985(b).
- 20 Nabors *et al.*, "Immunization of healthy adults with a single recombinant pneumococcal surface protein A (PspA) variant stimulates broadly cross-reactive antibodies to heterologous PspA molecules," *Vaccine* 18:1743-1754, 2000.
- Nakai and Kanehisa, "Expert system for predicting protein localization sites in gram-negative bacteria," *Proteins* 11(2):95-110, 1991.
- 25 Navarre and Schneewind, "Surface Proteins of Gram-Positive Bacteria and Mechanisms of Their Targeting to the Cell Wall Envelope," *Microbiol. Mol. Biol. Rev.* 63(1):174-229, 1999.
- Nielsen *et al.*, "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites," *Protein Engineering* 10(1):1-6, 1997.
- 30 O'Gon-nan *et al.*, *Science* 251:1351-1355, 1991.
- Olmsted *et al.*, "High-resolution visualization by field emission scanning electron microscopy of *Enterococcus faecalis* surface proteins encoded by the



- pheromone-inducible conjugative plasmid pCF10," *J. Bacteriol.* 175(19):6229-37, 1993.
- Orita *et al.*, *PNAS* 86:2766, 1989.
- Orihuela *et al.*, "Peritoneal culture alters *Streptococcus pneumoniae* protein profiles and virulence properties," *Infect. Immun.* 68:6082-6086, 2000.
- 5 Park and Teichmann, "DIVCLUS: an automatic method in the GEANFAMMER package that finds homologous domains in single- and multi-domain proteins," *Bioinformatics* 14(2):144-50, 1998.
- Parkhill *et al.*, "Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491," *Nature* 404(6777):502-6, 2000.
- 10 Pierschbacher and Ruoslahti, "Influence of stereochemistry of the sequence Arg-Gly-Asp-Xaa on binding specificity in cell adhesion," *J. Biol. Chem.* 262(36):17294-8, 1987.
- Pinkert *et al.* *Genes Dev.* 1:268-277, 1987.
- 15 Pizza *et al.*, "Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing," *Science* 287(5459):1816-20, 2000.
- Pugsley, "The complete general secretory pathway in gram-negative bacteria," *Microbiol. Rev.* 57(1):50-108, 1993.
- Queen and Baltimore, *Cell* 33:741-748, 1983.
- 20 Rahman *et al.*, *Journal of Neuroscience* 19:2016-2026, 1999.
- Rose *et al.*, "Methods in Yeast Genetics: A Laboratory Course Manual." *Cold Spring Harbor Press*, Cold Spring Harbor, N.Y. (1990).
- Rosenow *et al.*, "Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of *Streptococcus pneumoniae*," *Mol. Microbiol.* 25(5):819-29, 1997.
- 25 Ross and Wilkie, "GTPase-activating proteins for Heterotrimeric G proteins: Regulators of G protein Signaling (RGS) and RGS-like proteins," *Annual Review of Biochemistry* 69:795-827, 2000.
- Saieeba *et al.*, *Meth. Enzymol.* 217:286-295, 1992.
- 30 Salzberg *et al.*, "Microbial gene identification using interpolated Markov models," *Nuc. Acids Res.* 26(2):544-8, 1998.

- Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual" 2nd, ed, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
- 5 Sampson *et al.*, "Cloning and nucleotide sequence analysis of *psaA*, the *Streptococcus pneumoniae* gene encoding a 37-kilodalton protein homologous to previously reported *Streptococcus* sp. Adhesins," *Infect. Immun.* 62(1):319-24, 1994.
- Sanger, PNAS 74:5463, 1977.
- Schultz *et al.*, *Gene* 54:113-123, 1987.
- 10 Seed, *Nature* 329:840, 1987.
- Shinefield and Black, "Efficacy of pneumococcal conjugate vaccines in large scale field trials (In Process Citation)," *Pediatr. Infect. Dis. J.* 19(4):394-7, 2000.
- Simon *et al.*, *Science*, 252:802-8, 1991.
- Smith and Johnson, *Gene* 67:31-40, 1988.
- 15 Smith *et al.*, *Mol. Cell Biol.* 3:2156-2165, 1983.
- Songyang, *et al.*, *Cell* 72:767-778, 1993.
- Sonnenberg and Belisle, "Definition of *Mycobacterium tuberculosis* culture filtrate proteins by two-dimensional polyacrylamide gel electrophoresis, N-terminal amino acid sequencing, and electrospray mass spectrometry," *Infect. Immun.* 65(11):4515-24, 1997.
- 20 Sonnhhammer *et al.*, "A hidden Markov model for predicting transmembrane helices in protein sequences," *Ismb* 6:175-82, 1998.
- Stockbauer *et al.*, "A natural variant of the cysteine protease virulence factor of group A streptococcus with an arginine-glycine-aspartic acid (RGD) motif preferentially binds human integrins  $\alpha$ 5 $\beta$ 1 and  $\alpha$ 5 $\beta$ 3 (In Process Citation)," *Proc. Natl. Acad. Sci. U S A* 96(1):242-7, 1999.
- 25 Studier *et al.* "Gene Expression Technology" *Methods in Enzymology* 185, 60-89, 1990.
- Talkington *et al.*, "Protection of mice against fatal pneumococcal challenge by immunization with pneumococcal surface adhesin A (*PsaA*)," *Microb. Pathog.* 21(1):17-22, 1996.
- 30

- Tebbey *et al.*, "Effective mucosal immunization against respiratory syncytial virus using a genetically detoxified cholera holotoxin, CT-E29H," *Vaccine* 18(24):2723-34, 2000.
- Thomas and Capecchi, *Cell* 51:503, 1987.
- 5 Weldingh *et al.*, "Two-dimensional electrophoresis for analysis of Mycobacterium tuberculosis culture filtrate and purification and characterization of six novel proteins," *Infect. Immun.* 66(8):3492-500, 1998.
- Wilmot *et al.*, *Nature* 385:810-813, 1997.
- Wilson *et al.*, *Cell* 37:767, 1984.
- 10 Winoto and Baltimore. *EMBO J* 8:729-733, 1989.
- Xu *et al.*, "PHR1 encodes an abundant, pleckstrin homology domain-containing Integral membrane protein in the photoreceptor outer segments," *Journal of Biological Chemistry* 274:35676-35685, 1999.
- Yamamoto *et al.*, "A nontoxic adjuvant for mucosal immunity to pneumococcal surface protein," *A. J. Immunol.* 161(8):4115-21, 1998.
- 15 Zervos *et al.*, *Cell* 72:223-232, 1993.
- Zhang *et al.*, 2001, "Recombinant PhpA Protein, a Unique Histidine Motif-Containing Protein from *Streptococcus pneumoniae*, Protects Mice against Intranasal Pneumococcal Challenge," *Infect. Immun.* 69:3827-3836, 2001.

What is Claimed is:

1. An isolated polynucleotide of a *Streptococcus pneumoniae* genomic sequence, wherein the polynucleotide comprises a nucleotide sequence having at least about 95% identity to a nucleotide sequence chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO:431 through SEQ ID NO:591, a degenerate variant thereof, or a fragment thereof.
2. The polynucleotide of claim 1, wherein the polynucleotide is a complement to a nucleotide sequence chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO:431 through SEQ ID NO:591, a degenerate variant thereof, or a fragment thereof.
3. The polynucleotide of claim 2, wherein the polynucleotide is selected from the group consisting of DNA, chromosomal DNA, cDNA and RNA.
4. The polynucleotide of claim 3, wherein the polynucleotide further comprises heterologous nucleotides.
5. An isolated polynucleotide which hybridizes to a nucleotide sequence chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO:431 through SEQ ID NO:591, a complement thereof, a degenerate variant thereof, or a fragment thereof, under high stringency hybridization conditions.
6. The polynucleotide of claim 5, wherein the polynucleotide hybridizes under intermediate stringency hybridization conditions.
7. An isolated polynucleotide of a *Streptococcus pneumoniae* genomic sequence, wherein the polynucleotide comprises a nucleotide sequence chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO:431 through SEQ ID NO:591, a fragment thereof, or a degenerate variant

thereof, and encodes a polypeptide, a biological equivalent thereof, or a fragment thereof, selected from the group consisting of:

- (a) a *Streptococcus pneumoniae* polypeptide having 0, 1 or 2 transmembrane domains;
  - (b) a *Streptococcus pneumoniae* polypeptide having 3 or more transmembrane domains;
  - (c) a *Streptococcus pneumoniae* polypeptide having an outer membrane domain or a periplasmic domain;
  - (d) a *Streptococcus pneumoniae* polypeptide having an inner membrane domain;
  - (e) a *Streptococcus pneumoniae* polypeptide identified by Blastp analysis;
  - (f) a *Streptococcus pneumoniae* polypeptide identified by Pfam analysis;
  - (g) a *Streptococcus pneumoniae* lipoprotein;
  - (h) a *Streptococcus pneumoniae* polypeptide having a LPXTG motif, wherein the polypeptide is covalently attached to the peptidoglycan layer;
  - (i) a *Streptococcus pneumoniae* polypeptide having a peptidoglycan binding motif, wherein the polypeptide is associated with the peptidoglycan layer;
  - (j) a *Streptococcus pneumoniae* polypeptide having a signal sequence and a C-terminal Tyrosine or a C-terminal Phenylalanine amino acid;
  - (k) a *Streptococcus pneumoniae* polypeptide having a tripeptide RGD amino acid sequence;
  - (l) a *Streptococcus pneumoniae* polypeptide identified by proteomics as surface exposed;
- and
- (m) a *Streptococcus pneumoniae* polypeptide identified by proteomics as membrane associated.

8. The polynucleotide of claim 7, wherein the polynucleotide is a complement to a nucleotide sequence chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO:431 through SEQ ID NO:591, a degenerate variant thereof, or a fragment thereof.

9. The polynucleotide of claim 8, wherein the polynucleotide is selected from the group consisting of DNA, chromosomal DNA, cDNA and RNA.
10. The polynucleotide of claim 9, wherein the polynucleotide further comprises heterologous nucleotides.
11. The polynucleotide of claim 10, wherein the polynucleotide encodes a fusion polypeptide.
12. The polynucleotide of claim 7, wherein the polynucleotide encoding a polypeptide having 0, 1 or 2 transmembrane domains comprises a nucleotide sequence chosen from one of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 39, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 45, SEQ ID NO: 47, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 53, SEQ ID NO: 55, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70, SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 78, SEQ ID NO: 79, SEQ ID NO: 81, SEQ ID NO: 83, SEQ ID NO: 85, SEQ ID NO: 86, SEQ ID NO: 89, SEQ ID NO: 91, SEQ ID NO: 92, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 97, SEQ ID NO: 100, SEQ ID NO: 104, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 109, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 113, SEQ ID NO: 116, SEQ ID NO: 121, SEQ ID NO: 122, SEQ ID NO: 123, SEQ ID NO: 125, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 128, SEQ ID NO: 131, SEQ ID NO: 132, SEQ ID NO: 134, SEQ ID NO: 136, SEQ ID NO: 137, SEQ ID NO: 138, SEQ ID NO: 141, SEQ ID NO: 142, SEQ ID NO: 143, SEQ ID NO: 144, SEQ ID NO: 147, SEQ ID NO: 148, SEQ ID NO: 149, SEQ ID NO: 150, SEQ ID NO: 155, SEQ ID NO: 156, SEQ ID NO: 158, SEQ ID

NO: 161, SEQ ID NO: 162, SEQ ID NO: 165, SEQ ID NO: 170, SEQ ID NO: 171, SEQ ID NO: 172, SEQ ID NO: 174, SEQ ID NO: 176, SEQ ID NO: 179, SEQ ID NO: 183, SEQ ID NO: 185, SEQ ID NO: 187, SEQ ID NO: 192, SEQ ID NO: 195, SEQ ID NO: 196, SEQ ID NO: 197, SEQ ID NO: 199, SEQ ID NO: 200, SEQ ID NO: 201, SEQ ID NO: 202, SEQ ID NO: 204, SEQ ID NO: 205, SEQ ID NO: 207, SEQ ID NO: 209 and SEQ ID NO: 210.

13. The polynucleotide of claim 7, wherein the polynucleotide encoding a polypeptide having 3 or more transmembrane domains comprises a nucleotide sequence chosen from one of SEQ ID NO: 2, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, SEQ ID NO: 52, SEQ ID NO: 54, SEQ ID NO: 56, SEQ ID NO: 59, SEQ ID NO: 65, SEQ ID NO: 71, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 80, SEQ ID NO: 82, SEQ ID NO: 84, SEQ ID NO: 87, SEQ ID NO: 88, SEQ ID NO: 90, SEQ ID NO: 93, SEQ ID NO: 94, SEQ ID NO: 98, SEQ ID NO: 99, SEQ ID NO: 101, SEQ ID NO: 102, SEQ ID NO: 103, SEQ ID NO: 107, SEQ ID NO: 108, SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 117, SEQ ID NO: 118, SEQ ID NO: 119, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 129, SEQ ID NO: 130, SEQ ID NO: 133, SEQ ID NO: 135, SEQ ID NO: 139, SEQ ID NO: 140, SEQ ID NO: 145, SEQ ID NO: 146, SEQ ID NO: 151, SEQ ID NO: 152, SEQ ID NO: 153, SEQ ID NO: 154, SEQ ID NO: 157, SEQ ID NO: 159, SEQ ID NO: 160, SEQ ID NO: 163, SEQ ID NO: 164, SEQ ID NO: 166, SEQ ID NO: 167, SEQ ID NO: 168, SEQ ID NO: 169, SEQ ID NO: 173, SEQ ID NO: 175, SEQ ID NO: 177, SEQ ID NO: 178, SEQ ID NO: 180, SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 184, SEQ ID NO: 186, SEQ ID NO: 188, SEQ ID NO: 189, SEQ ID NO: 190, SEQ ID NO: 191, SEQ ID NO: 193, SEQ ID NO: 194, SEQ ID NO: 198, SEQ ID NO: 203, SEQ ID NO: 206, SEQ ID NO: 208, SEQ ID NO: 211, SEQ ID NO: 212, SEQ ID NO: 213, SEQ ID NO: 214 and SEQ ID NO: 215.

14. The polynucleotide of claim 7, wherein the polynucleotide encoding a polypeptide having an outer membrane domain or a periplasmic domain comprises a nucleotide sequence chosen from one of SEQ ID NO: 3, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 23, SEQ ID NO: 39, SEQ ID NO: 50, SEQ ID NO: 62, SEQ ID NO: 67, SEQ ID NO: 78, SEQ ID NO: 85, SEQ ID NO: 125, SEQ ID NO: 134, SEQ ID NO: 147, SEQ ID NO: 165, SEQ ID NO: 172 and SEQ ID NO: 179.
15. The polynucleotide of claim 7, wherein the polynucleotide encoding a polypeptide having an inner membrane domain comprises a nucleotide sequence chosen from one of SEQ ID NO: 2, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 56, SEQ ID NO: 59, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 65, SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70, SEQ ID NO: 71, SEQ ID NO: 73, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 81, SEQ ID NO: 82, SEQ ID NO: 83, SEQ ID NO: 84, SEQ ID NO: 86, SEQ ID NO: 87, SEQ ID NO: 88, SEQ ID NO: 90, SEQ ID NO: 91, SEQ ID NO: 93, SEQ ID NO: 94, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 97, SEQ ID NO: 98, SEQ ID NO: 99, SEQ ID NO: 100, SEQ ID NO: 101, SEQ ID NO: 102, SEQ ID NO: 103, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 107, SEQ ID NO: 108, SEQ ID NO: 109, SEQ ID NO: 112, SEQ ID NO: 113, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 117, SEQ ID NO: 118, SEQ ID NO: 119, SEQ ID NO: 120, SEQ ID NO: 121, SEQ ID NO: 122, SEQ ID NO: 123, SEQ ID NO: 124, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 130, SEQ ID



NO: 131, SEQ ID NO: 132, SEQ ID NO: 133, SEQ ID NO: 135, SEQ ID NO: 136, SEQ ID NO: 139, SEQ ID NO: 140, SEQ ID NO: 141, SEQ ID NO: 142, SEQ ID NO: 144, SEQ ID NO: 145, SEQ ID NO: 146, SEQ ID NO: 148, SEQ ID NO: 150, SEQ ID NO: 151, SEQ ID NO: 152, SEQ ID NO: 153, SEQ ID NO: 154, SEQ ID NO: 156, SEQ ID NO: 157, SEQ ID NO: 158, SEQ ID NO: 159, SEQ ID NO: 160, SEQ ID NO: 162, SEQ ID NO: 163, SEQ ID NO: 164, SEQ ID NO: 166, SEQ ID NO: 167, SEQ ID NO: 168, SEQ ID NO: 169, SEQ ID NO: 170, SEQ ID NO: 173, SEQ ID NO: 175, SEQ ID NO: 176, SEQ ID NO: 177, SEQ ID NO: 178, SEQ ID NO: 180, SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 184, SEQ ID NO: 186, SEQ ID NO: 187, SEQ ID NO: 188, SEQ ID NO: 189, SEQ ID NO: 190, SEQ ID NO: 191, SEQ ID NO: 192, SEQ ID NO: 193, SEQ ID NO: 194, SEQ ID NO: 195, SEQ ID NO: 198, SEQ ID NO: 200, SEQ ID NO: 203, SEQ ID NO: 206, SEQ ID NO: 208, SEQ ID NO: 209, SEQ ID NO: 211, SEQ ID NO: 212, SEQ ID NO: 213, SEQ ID NO: 214 and SEQ ID NO: 215.

16. The polynucleotide of claim 7, wherein the polynucleotide encoding a polypeptide identified by Blastp analysis comprises a nucleotide sequence chosen from one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 7, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 16, SEQ ID NO: 20, SEQ ID NO: 24, SEQ ID NO: 27, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 48, SEQ ID NO: 51, SEQ ID NO: 53, SEQ ID NO: 59, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 65, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70, SEQ ID NO: 71, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 87, SEQ ID NO: 88, SEQ ID NO: 90, SEQ ID NO: 94, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 98, SEQ ID NO: 100, SEQ ID NO: 103, SEQ ID NO: 105, SEQ ID NO: 107, SEQ ID NO: 108, SEQ ID NO: 109, SEQ ID NO: 112, SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 117, SEQ ID NO: 118, SEQ ID NO: 122, SEQ ID NO: 123, SEQ ID NO: 124, SEQ ID NO: 127, SEQ ID NO: 129, SEQ ID NO: 131, SEQ ID NO: 132, SEQ ID NO: 133, SEQ ID NO: 134, SEQ

ID NO: 135, SEQ ID NO: 136, SEQ ID NO: 138, SEQ ID NO: 139, SEQ ID NO: 141, SEQ ID NO: 144, SEQ ID NO: 146, SEQ ID NO: 147, SEQ ID NO: 151, SEQ ID NO: 152, SEQ ID NO: 154, SEQ ID NO: 155, SEQ ID NO: 157, SEQ ID NO: 158, SEQ ID NO: 159, SEQ ID NO: 160, SEQ ID NO: 161, SEQ ID NO: 162, SEQ ID NO: 163, SEQ ID NO: 165, SEQ ID NO: 166, SEQ ID NO: 167, SEQ ID NO: 169, SEQ ID NO: 172, SEQ ID NO: 173, SEQ ID NO: 176, SEQ ID NO: 177, SEQ ID NO: 178, SEQ ID NO: 180, SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 184, SEQ ID NO: 185, SEQ ID NO: 186, SEQ ID NO: 188, SEQ ID NO: 189, SEQ ID NO: 191, SEQ ID NO: 193, SEQ ID NO: 196, SEQ ID NO: 197, SEQ ID NO: 198, SEQ ID NO: 199, SEQ ID NO: 200, SEQ ID NO: 201, SEQ ID NO: 202, SEQ ID NO: 204, SEQ ID NO: 205, SEQ ID NO: 206, SEQ ID NO: 207, SEQ ID NO: 208, SEQ ID NO: 210, SEQ ID NO: 212, SEQ ID NO: 213 and SEQ ID NO: 214.

17. The polynucleotide of claim 7, wherein the polynucleotide encoding a polypeptide identified by Pfam analysis comprises a nucleotide sequence chosen from one of SEQ ID NO: 4, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 41, SEQ ID NO: 45, SEQ ID NO: 55, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 66, SEQ ID NO: 72, SEQ ID NO: 74, SEQ ID NO: 89, SEQ ID NO: 92, SEQ ID NO: 104, SEQ ID NO: 111, SEQ ID NO: 116, SEQ ID NO: 119, SEQ ID NO: 128, SEQ ID NO: 137, SEQ ID NO: 142, SEQ ID NO: 143, SEQ ID NO: 149, SEQ ID NO: 151, SEQ ID NO: 152, SEQ ID NO: 153, SEQ ID NO: 157, SEQ ID NO: 159, SEQ ID NO: 160, SEQ ID NO: 162, SEQ ID NO: 163, SEQ ID NO: 164, SEQ ID NO: 165, SEQ ID NO: 166, SEQ ID NO: 169, SEQ ID NO: 171, SEQ ID NO: 174, SEQ ID NO: 176, SEQ ID NO: 180, SEQ ID NO: 182, SEQ ID NO: 183, SEQ ID NO: 184, SEQ ID NO: 186, SEQ ID NO: 188, SEQ ID NO: 189, SEQ ID NO: 195, SEQ ID NO: 198, SEQ ID NO: 199, SEQ ID NO: 205, SEQ ID NO: 212 and SEQ ID NO: 213.
18. The polynucleotide of claim 7, wherein the polynucleotide encoding a lipoprotein comprises a nucleotide sequence chosen from one of SEQ ID NO: 3, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 13, SEQ ID NO: 21, SEQ ID

NO: 26, SEQ ID NO: 34, SEQ ID NO: 62, SEQ ID NO: 67, SEQ ID NO: 85, SEQ ID NO: 134, SEQ ID NO: 147, SEQ ID NO: 150, SEQ ID NO: 168, SEQ ID NO: 170 and SEQ ID NO: 173.

19. The polynucleotide of claim 7, wherein the polynucleotide encoding a polypeptide having a LPXTG motif and covalently attached to the peptidoglycan layer comprises a nucleotide sequence chosen from one of SEQ ID NO: 13, SEQ ID NO: 21, SEQ ID NO: 34 and SEQ ID NO: 170.
20. The polynucleotide of claim 7, wherein the polynucleotide encoding a polypeptide having a peptidoglycan binding motif and associated with the peptidoglycan layer comprises a nucleotide sequence chosen from one of SEQ ID NO: 25, SEQ ID NO: 49 and SEQ ID NO: 110.
21. The polynucleotide of claim 7, wherein the polynucleotide encoding a polypeptide having a signal sequence and a C-terminal Tyrosine or a C-terminal Phenylalanine amino acid comprises a nucleotide sequence chosen from one of SEQ ID NO:11, SEQ ID NO:39, SEQ ID NO:73, SEQ ID NO:97, SEQ ID NO:106, SEQ ID NO: 125 and SEQ ID NO:187.
22. The polynucleotide of claim 7, wherein the polynucleotide encoding a polypeptide having a tripeptide RGD amino acid sequence comprises a nucleotide sequence chosen from one of SEQ ID NO:1, SEQ ID NO:21, SEQ ID NO:66 and SEQ ID NO:67.
23. The polynucleotide of claim 7, wherein the polynucleotide encoding a polypeptide identified by proteomics as surface exposed comprises a nucleotide sequence chosen from one of SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:46, SEQ ID NO:64, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:69, SEQ ID NO:71, SEQ ID NO:74, SEQ ID NO:91, SEQ ID NO:103, SEQ ID NO:116, SEQ ID NO:128, SEQ ID NO:131, SEQ ID NO:136, SEQ ID NO:151, SEQ ID NO:156, SEQ ID NO:159, SEQ ID NO:162,

SEQ ID NO:164, SEQ ID NO:172, SEQ ID NO:176, SEQ ID NO:178, SEQ ID NO:179, SEQ ID NO:180, SEQ ID NO:182 and SEQ ID NO:205.

24. The polynucleotide of claim 7, wherein the polynucleotide encoding a polypeptide identified by proteomics as membrane associated comprises a nucleotide sequence chosen from one of SEQ ID NO:431 through SEQ ID NO: 591.
25. An isolated polypeptide encoded by a polynucleotide of a *Streptococcus pneumoniae* genomic sequence, wherein the polynucleotide comprises a nucleotide sequence having at least about 95% identity to a nucleotide sequence chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO: 431 through SEQ ID NO: 591, a degenerate variant thereof, or a fragment thereof.
26. The polypeptide of claim 25, wherein the polypeptide is a fusion polypeptide.
27. The polypeptide of claim 25, which immunoreacts with seropositive serum of an individual infected with *Streptococcus pneumoniae*.
28. The polypeptide of claim 25, further defined as:
  - (a) a *Streptococcus pneumoniae* polypeptide having 0, 1 or 2 transmembrane domains;
  - (b) a *Streptococcus pneumoniae* polypeptide having 3 or more transmembrane domains;
  - (c) a *Streptococcus pneumoniae* polypeptide having an outer membrane domain or a periplasmic domain;
  - (d) a *Streptococcus pneumoniae* polypeptide having an inner membrane domain;
  - (e) a *Streptococcus pneumoniae* polypeptide identified by Blastp analysis;
  - (f) a *Streptococcus pneumoniae* polypeptide identified by Pfam analysis;
  - (g) a *Streptococcus pneumoniae* lipoprotein;

- (h) a *Streptococcus pneumoniae* polypeptide having a LPXTG motif, wherein the polypeptide is covalently attached to the peptidoglycan layer;
  - (i) a *Streptococcus pneumoniae* polypeptide having a peptidoglycan binding motif, wherein the polypeptide is associated with the peptidoglycan layer;
  - (j) a *Streptococcus pneumoniae* polypeptide having a signal sequence and a C-terminal Tyrosine or a C-terminal Phenylalanine amino acid;
  - (k) a *Streptococcus pneumoniae* polypeptide having a tripeptide RGD amino acid sequence;
  - (l) a *Streptococcus pneumoniae* polypeptide identified by proteomics as surface exposed;
- and
- (m) a *Streptococcus pneumoniae* polypeptide identified by proteomics as membrane associated.

29. The polypeptide of claim 28, wherein the polypeptide having 0, 1 or 2 transmembrane domains comprises an amino acid sequence chosen from one of SEQ ID NO: 216, SEQ ID NO: 218, SEQ ID NO: 219, SEQ ID NO: 222, SEQ ID NO: 223, SEQ ID NO: 224, SEQ ID NO: 226, SEQ ID NO: 228, SEQ ID NO: 231, SEQ ID NO: 232, SEQ ID NO: 233, SEQ ID NO: 234, SEQ ID NO: 237, SEQ ID NO: 238, SEQ ID NO: 239, SEQ ID NO: 240, SEQ ID NO: 243, SEQ ID NO: 244, SEQ ID NO: 247, SEQ ID NO: 249, SEQ ID NO: 251, SEQ ID NO: 254, SEQ ID NO: 256, SEQ ID NO: 257, SEQ ID NO: 260, SEQ ID NO: 262, SEQ ID NO: 264, SEQ ID NO: 265, SEQ ID NO: 266, SEQ ID NO: 268, SEQ ID NO: 270, SEQ ID NO: 272, SEQ ID NO: 273, SEQ ID NO: 275, SEQ ID NO: 276, SEQ ID NO: 277, SEQ ID NO: 278, SEQ ID NO: 279, SEQ ID NO: 281, SEQ ID NO: 282, SEQ ID NO: 283, SEQ ID NO: 284, SEQ ID NO: 285, SEQ ID NO: 286, SEQ ID NO: 287, SEQ ID NO: 289, SEQ ID NO: 293, SEQ ID NO: 294, SEQ ID NO: 296, SEQ ID NO: 298, SEQ ID NO: 300, SEQ ID NO: 301, SEQ ID NO: 304, SEQ ID NO: 306, SEQ ID NO: 307, SEQ ID NO: 310, SEQ ID NO: 311, SEQ ID NO: 312, SEQ ID NO: 315, SEQ ID NO: 319, SEQ ID NO: 320, SEQ ID NO: 321, SEQ ID NO: 324, SEQ

ID NO: 325, SEQ ID NO: 326, SEQ ID NO: 328, SEQ ID NO: 331, SEQ ID NO: 336, SEQ ID NO: 337, SEQ ID NO: 338, SEQ ID NO: 340, SEQ ID NO: 341, SEQ ID NO: 342, SEQ ID NO: 343, SEQ ID NO: 346, SEQ ID NO: 347, SEQ ID NO: 349, SEQ ID NO: 351, SEQ ID NO: 352, SEQ ID NO: 353, SEQ ID NO: 356, SEQ ID NO: 357, SEQ ID NO: 358, SEQ ID NO: 359, SEQ ID NO: 362, SEQ ID NO: 363, SEQ ID NO: 364, SEQ ID NO: 365, SEQ ID NO: 370, SEQ ID NO: 371, SEQ ID NO: 373, SEQ ID NO: 376, SEQ ID NO: 377, SEQ ID NO: 380, SEQ ID NO: 385, SEQ ID NO: 386, SEQ ID NO: 387, SEQ ID NO: 389, SEQ ID NO: 391, SEQ ID NO: 394, SEQ ID NO: 398, SEQ ID NO: 400, SEQ ID NO: 402, SEQ ID NO: 407, SEQ ID NO: 410, SEQ ID NO: 411, SEQ ID NO: 412, SEQ ID NO: 414, SEQ ID NO: 415, SEQ ID NO: 416, SEQ ID NO: 417, SEQ ID NO: 419, SEQ ID NO: 420, SEQ ID NO: 422, SEQ ID NO: 424, SEQ ID NO: 425, a fragment thereof or a degenerate variant thereof.

30. The polypeptide of claim 28, wherein the polypeptide having 3 or more transmembrane domains comprises an amino acid sequence chosen from one of SEQ ID NO: 217, SEQ ID NO: 220, SEQ ID NO: 221, SEQ ID NO: 225, SEQ ID NO: 227, SEQ ID NO: 229, SEQ ID NO: 230, SEQ ID NO: 235, SEQ ID NO: 236, SEQ ID NO: 241, SEQ ID NO: 242, SEQ ID NO: 245, SEQ ID NO: 246, SEQ ID NO: 248, SEQ ID NO: 250, SEQ ID NO: 252, SEQ ID NO: 253, SEQ ID NO: 255, SEQ ID NO: 258, SEQ ID NO: 259, SEQ ID NO: 261, SEQ ID NO: 263, SEQ ID NO: 267, SEQ ID NO: 269, SEQ ID NO: 271, SEQ ID NO: 274, SEQ ID NO: 280, SEQ ID NO: 286, SEQ ID NO: 290, SEQ ID NO: 291, SEQ ID NO: 292, SEQ ID NO: 295, SEQ ID NO: 297, SEQ ID NO: 299, SEQ ID NO: 302, SEQ ID NO: 303, SEQ ID NO: 305, SEQ ID NO: 308, SEQ ID NO: 309, SEQ ID NO: 313, SEQ ID NO: 314, SEQ ID NO: 316, SEQ ID NO: 317, SEQ ID NO: 318, SEQ ID NO: 322, SEQ ID NO: 323, SEQ ID NO: 327, SEQ ID NO: 329, SEQ ID NO: 330, SEQ ID NO: 332, SEQ ID NO: 333, SEQ ID NO: 334, SEQ ID NO: 335, SEQ ID NO: 339, SEQ ID NO: 344, SEQ ID NO: 345, SEQ ID NO: 348, SEQ ID NO: 350, SEQ ID NO: 354, SEQ ID NO: 355, SEQ ID NO: 360, SEQ ID NO: 361, SEQ ID NO: 366, SEQ ID NO: 367, SEQ ID NO: 368, SEQ ID NO: 369, SEQ ID NO: 372, SEQ ID

NO: 374, SEQ ID NO: 375, SEQ ID NO: 378, SEQ ID NO: 379, SEQ ID NO: 381, SEQ ID NO: 382, SEQ ID NO: 383, SEQ ID NO: 384, SEQ ID NO: 388, SEQ ID NO: 390, SEQ ID NO: 392, SEQ ID NO: 393, SEQ ID NO: 395, SEQ ID NO: 396, SEQ ID NO: 397, SEQ ID NO: 399, SEQ ID NO: 401, SEQ ID NO: 403, SEQ ID NO: 404, SEQ ID NO: 405, SEQ ID NO: 406, SEQ ID NO: 408, SEQ ID NO: 409, SEQ ID NO: 413, SEQ ID NO: 418, SEQ ID NO: 421, SEQ ID NO: 423, SEQ ID NO: 426, SEQ ID NO: 427, SEQ ID NO: 428, SEQ ID NO: 429, SEQ ID NO: 430, a fragment thereof or a degenerate variant thereof.

31. The polypeptide of claim 28, wherein the polypeptide having an outer membrane or a periplasmic domain comprises an amino acid sequence chosen from one of SEQ ID NO: 218, SEQ ID NO: 223, SEQ ID NO: 224, SEQ ID NO: 238, SEQ ID NO: 254, SEQ ID NO: 265, SEQ ID NO: 277, SEQ ID NO: 282, SEQ ID NO: 293, SEQ ID NO: 300, SEQ ID NO: 340, SEQ ID NO: 349, SEQ ID NO: 362, SEQ ID NO: 380, SEQ ID NO: 387, SEQ ID NO: 394, a fragment thereof or a degenerate variant thereof.
32. The polypeptide of claim 28, wherein the polypeptide having an inner membrane domain comprises an amino acid sequence chosen from one of SEQ ID NO: 217, SEQ ID NO: 220, SEQ ID NO: 221, SEQ ID NO: 222, SEQ ID NO: 225, SEQ ID NO: 226, SEQ ID NO: 227, SEQ ID NO: 228, SEQ ID NO: 229, SEQ ID NO: 230, SEQ ID NO: 231, SEQ ID NO: 232, SEQ ID NO: 234, SEQ ID NO: 235, SEQ ID NO: 236, SEQ ID NO: 237, SEQ ID NO: 241, SEQ ID NO: 242, SEQ ID NO: 243, SEQ ID NO: 244, SEQ ID NO: 245, SEQ ID NO: 246, SEQ ID NO: 247, SEQ ID NO: 248, SEQ ID NO: 249, SEQ ID NO: 250, SEQ ID NO: 251, SEQ ID NO: 252, SEQ ID NO: 253, SEQ ID NO: 255, SEQ ID NO: 258, SEQ ID NO: 259, SEQ ID NO: 261, SEQ ID NO: 262, SEQ ID NO: 263, SEQ ID NO: 266, SEQ ID NO: 267, SEQ ID NO: 268, SEQ ID NO: 269, SEQ ID NO: 271, SEQ ID NO: 274, SEQ ID NO: 275, SEQ ID NO: 276, SEQ ID NO: 280, SEQ ID NO: 283, SEQ ID NO: 284, SEQ ID NO: 285, SEQ ID NO: 286, SEQ ID NO: 288, SEQ ID NO: 290, SEQ ID NO: 291, SEQ ID NO: 292, SEQ ID NO: 294, SEQ ID NO: 295, SEQ ID NO: 296, SEQ

ID NO: 297, SEQ ID NO: 298, SEQ ID NO: 299, SEQ ID NO: 301 SEQ ID NO: 302, SEQ ID NO: 303, SEQ ID NO: 305, SEQ ID NO: 306, SEQ ID NO: 308, SEQ ID NO: 309, SEQ ID NO: 310, SEQ ID NO: 311, SEQ ID NO: 312, SEQ ID NO: 313, SEQ ID NO: 314, SEQ ID NO: 315, SEQ ID NO: 316, SEQ ID NO: 317, SEQ ID NO: 318, SEQ ID NO: 320, SEQ ID NO: 321, SEQ ID NO: 322, SEQ ID NO: 323, SEQ ID NO: 324, SEQ ID NO: 327, SEQ ID NO: 328, SEQ ID NO: 329, SEQ ID NO: 330, SEQ ID NO: 332, SEQ ID NO: 333, SEQ ID NO: 334, SEQ ID NO: 335, SEQ ID NO: 336, SEQ ID NO: 337, SEQ ID NO: 338, SEQ ID NO: 339, SEQ ID NO: 341, SEQ ID NO: 342, SEQ ID NO: 343, SEQ ID NO: 344, SEQ ID NO: 345, SEQ ID NO: 346, SEQ ID NO: 347, SEQ ID NO: 348, SEQ ID NO: 350, SEQ ID NO: 351, SEQ ID NO: 354, SEQ ID NO: 355, SEQ ID NO: 356, SEQ ID NO: 357, SEQ ID NO: 359, SEQ ID NO: 360, SEQ ID NO: 361, SEQ ID NO: 362, SEQ ID NO: 365, SEQ ID NO: 366, SEQ ID NO: 367, SEQ ID NO: 368, SEQ ID NO: 369, SEQ ID NO: 371, SEQ ID NO: 372, SEQ ID NO: 373, SEQ ID NO: 374, SEQ ID NO: 375, SEQ ID NO: 377, SEQ ID NO: 378, SEQ ID NO: 379, SEQ ID NO: 381, SEQ ID NO: 382, SEQ ID NO: 383, SEQ ID NO: 384, SEQ ID NO: 385, SEQ ID NO: 388, SEQ ID NO: 390, SEQ ID NO: 391, SEQ ID NO: 392, SEQ ID NO: 393, SEQ ID NO: 395, SEQ ID NO: 396, SEQ ID NO: 397, SEQ ID NO: 399, SEQ ID NO: 401, SEQ ID NO: 402, SEQ ID NO: 403, SEQ ID NO: 404, SEQ ID NO: 405, SEQ ID NO: 406, SEQ ID NO: 407, SEQ ID NO: 408, SEQ ID NO: 409, SEQ ID NO: 410, SEQ ID NO: 413, SEQ ID NO: 415, SEQ ID NO: 418, SEQ ID NO: 421, SEQ ID NO: 423, SEQ ID NO: 424, SEQ ID NO: 426, SEQ ID NO: 427, SEQ ID NO: 428, SEQ ID NO: 429, SEQ ID NO: 430, a fragment thereof or a degenerate variant thereof.

33. The polypeptide of claim 28, wherein the polypeptide identified by Blastp analysis comprises an amino acid sequence chosen from one of SEQ ID NO: 216, SEQ ID NO: 217, SEQ ID NO: 222, SEQ ID NO: 225, SEQ ID NO: 227, SEQ ID NO: 231, SEQ ID NO: 235, SEQ ID NO: 239, SEQ ID NO: 242, SEQ ID NO: 245, SEQ ID NO: 246, SEQ ID NO: 247, SEQ ID NO: 248, SEQ ID NO: 249, SEQ ID NO: 250, SEQ ID NO: 253, SEQ ID NO: 255, SEQ ID NO: 257, SEQ ID NO: 258, SEQ ID NO: 259, SEQ ID NO: 263, SEQ ID NO: 266,



SEQ ID NO: 268, SEQ ID NO: 269, SEQ ID NO: 275, SEQ ID NO: 276, SEQ ID NO: 280, SEQ ID NO: 282, SEQ ID NO: 283, SEQ ID NO: 284, SEQ ID NO: 285, SEQ ID NO: 286, SEQ ID NO: 290, SEQ ID NO: 291, SEQ ID NO: 292, SEQ ID NO: 293, SEQ ID NO: 294, SEQ ID NO: 295, SEQ ID NO: 302, SEQ ID NO: 303, SEQ ID NO: 305, SEQ ID NO: 309, SEQ ID NO: 310, SEQ ID NO: 311, SEQ ID NO: 313, SEQ ID NO: 315, SEQ ID NO: 318, SEQ ID NO: 320, SEQ ID NO: 322, SEQ ID NO: 323, SEQ ID NO: 324, SEQ ID NO: 327, SEQ ID NO: 328, SEQ ID NO: 330, SEQ ID NO: 332, SEQ ID NO: 333, SEQ ID NO: 337, SEQ ID NO: 338, SEQ ID NO: 339, SEQ ID NO: 342, SEQ ID NO: 344, SEQ ID NO: 346, SEQ ID NO: 347, SEQ ID NO: 348, SEQ ID NO: 349, SEQ ID NO: 350, SEQ ID NO: 351, SEQ ID NO: 353, SEQ ID NO: 354, SEQ ID NO: 356, SEQ ID NO: 359, SEQ ID NO: 361, SEQ ID NO: 362, SEQ ID NO: 366, SEQ ID NO: 367, SEQ ID NO: 369, SEQ ID NO: 370, SEQ ID NO: 372, SEQ ID NO: 373, SEQ ID NO: 374, SEQ ID NO: 375, SEQ ID NO: 376, SEQ ID NO: 377, SEQ ID NO: 378, SEQ ID NO: 380, SEQ ID NO: 381, SEQ ID NO: 382, SEQ ID NO: 384, SEQ ID NO: 387, SEQ ID NO: 388, SEQ ID NO: 391, SEQ ID NO: 392, SEQ ID NO: 393, SEQ ID NO: 395, SEQ ID NO: 396, SEQ ID NO: 397, SEQ ID NO: 399, SEQ ID NO: 400, SEQ ID NO: 401, SEQ ID NO: 403, SEQ ID NO: 404, SEQ ID NO: 406, SEQ ID NO: 408, SEQ ID NO: 411, SEQ ID NO: 412, SEQ ID NO: 413, SEQ ID NO: 414, SEQ ID NO: 415, SEQ ID NO: 416, SEQ ID NO: 417, SEQ ID NO: 419, SEQ ID NO: 420, SEQ ID NO: 421, SEQ ID NO: 422, SEQ ID NO: 423, SEQ ID NO: 425, SEQ ID NO: 427, SEQ ID NO: 428, SEQ ID NO: 429, a fragment thereof or a degenerate variant thereof.

34. The polypeptide of claim 28, wherein the polypeptide identified by Pfam analysis comprises an amino acid sequence chosen from one of SEQ ID NO: 219, SEQ ID NO: 233, SEQ ID NO: 234, SEQ ID NO: 255, SEQ ID NO: 260, SEQ ID NO: 270, SEQ ID NO: 272, SEQ ID NO: 273, SEQ ID NO: 278, SEQ ID NO: 279, SEQ ID NO: 281, SEQ ID NO: 287, SEQ ID NO: 289, SEQ ID NO: 304, SEQ ID NO: 307, SEQ ID NO: 319, SEQ ID NO: 326, SEQ ID NO: 331, SEQ ID NO: 334, SEQ ID NO: 343, SEQ ID NO: 352, SEQ ID NO: 357, SEQ ID NO: 358, SEQ ID NO: 364, SEQ ID NO: 366, SEQ ID NO: 367, SEQ

- ID NO: 368, SEQ ID NO: 372, SEQ ID NO: 374, SEQ ID NO: 375, SEQ ID NO: 377, SEQ ID NO: 378, SEQ ID NO: 379, SEQ ID NO: 380, SEQ ID NO: 381, SEQ ID NO: 384, SEQ ID NO: 386, SEQ ID NO: 389, SEQ ID NO: 391, SEQ ID NO: 395, SEQ ID NO: 397, SEQ ID NO: 398, SEQ ID NO: 399, SEQ ID NO: 401, SEQ ID NO: 403, SEQ ID NO: 404, SEQ ID NO: 410, SEQ ID NO: 413, SEQ ID NO: 414, SEQ ID NO: 420, SEQ ID NO: 427, SEQ ID NO: 428, a fragment thereof or a degenerate variant thereof.
35. The polypeptide of claim 28, wherein the polypeptide is a lipoprotein and comprises an amino acid sequence chosen from one of SEQ ID NO: 218, SEQ ID NO: 223, SEQ ID NO: 224, SEQ ID NO: 228, SEQ ID NO: 236, SEQ ID NO: 241, SEQ ID NO: 249, SEQ ID NO: 277, SEQ ID NO: 282, SEQ ID NO: 300, SEQ ID NO: 349, SEQ ID NO: 362, SEQ ID NO: 365, SEQ ID NO: 383, SEQ ID NO: 385, SEQ ID NO: 388, a fragment thereof or a degenerate variant thereof.
36. The polypeptide of claim 28, wherein the polypeptide having a LPXTG motif and covalently attached to the peptidoglycan layer comprises an amino acid sequence chosen from one of SEQ ID NO: 228, SEQ ID NO: 236, SEQ ID NO: 249, SEQ ID NO: 385, a fragment thereof or a degenerate variant thereof.
37. The polypeptide of claim 28, wherein the polypeptide having a peptidoglycan binding motif and associated with the peptidoglycan layer comprises an amino acid sequence selected from one of SEQ ID NO: 240, SEQ ID NO: 264, SEQ ID NO: 325, a fragment thereof or a degenerate variant thereof.
38. The polypeptide of claim 28, wherein the polypeptide having a signal sequence and a C-terminal Tyrosine or Phenylalanine amino acid comprises an amino acid sequence chosen from one of SEQ ID NO: 226, SEQ ID NO: 254, SEQ ID NO: 289, SEQ ID NO: 312, SEQ ID NO: 321, SEQ ID NO: 340, SEQ ID NO: 402, a fragment thereof or a degenerate variant thereof.

39. The polypeptide of claim 28, wherein the polypeptide having a tripeptide RGD sequence that potentially is involved in cell attachment comprises an amino acid sequence chosen from one of SEQ ID NO:216, SEQ ID NO:236, SEQ ID NO:281, SEQ ID NO:282, a fragment thereof or a degenerate variant thereof.
40. The polypeptide of claim 28, wherein the polypeptide identified by proteomics as surface exposed comprises an amino acid sequence chosen from one of SEQ ID NO: 229, SEQ ID NO: 231, SEQ ID NO: 232, SEQ ID NO: 261, SEQ ID NO: 279, SEQ ID NO: 281, SEQ ID NO: 282, SEQ ID NO: 284, SEQ ID NO: 286, SEQ ID NO: 289, SEQ ID NO: 306, SEQ ID NO: 318, SEQ ID NO: 331, SEQ ID NO: 343, SEQ ID NO: 346, SEQ ID NO: 351, SEQ ID NO: 366, SEQ ID NO: 371, SEQ ID NO: 374, SEQ ID NO: 377, SEQ ID NO: 379, SEQ ID NO: 387, SEQ ID NO: 391, SEQ ID NO: 393, SEQ ID NO: 394, 395, SEQ ID NO: 397, SEQ ID NO: 420, a fragment thereof or a degenerate variant thereof.
41. The polypeptide of claim 28, wherein the polypeptide identified by proteomics as surface exposed comprises an amino acid sequence chosen from one of SEQ ID NO:592 through SEQ ID NO: 752, a fragment thereof or a degenerate variant thereof.
42. An isolated polypeptide comprising an amino acid sequence having at least about 95% identity to an amino acid sequence chosen from one of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID NO: 592 through SEQ ID NO:752.
43. The polypeptide of claim 42, wherein the polypeptide is a fusion polypeptide
44. The polypeptide of claim 42, which immunoreacts with seropositive serum of an individual infected with *Streptococcus pneumoniae*.
45. The polypeptide of claim 42, further defined as:

- (a) a *Streptococcus pneumoniae* polypeptide having 0, 1 or 2 transmembrane domains;
  - (b) a *Streptococcus pneumoniae* polypeptide having 3 or more transmembrane domains;
  - (c) a *Streptococcus pneumoniae* polypeptide having an outer membrane domain or a periplasmic domain;
  - (d) a *Streptococcus pneumoniae* polypeptide having an inner membrane domain;
  - (e) a *Streptococcus pneumoniae* polypeptide identified by Blastp analysis;
  - (f) a *Streptococcus pneumoniae* polypeptide identified by Pfam analysis;
  - (g) a *Streptococcus pneumoniae* lipoprotein;
  - (h) a *Streptococcus pneumoniae* polypeptide having a LPXTG motif, wherein the polypeptide is covalently attached to the peptidoglycan layer;
  - (i) a *Streptococcus pneumoniae* polypeptide having a peptidoglycan binding motif, wherein the polypeptide is associated with the peptidoglycan layer;
  - (j) a *Streptococcus pneumoniae* polypeptide having a signal sequence and a C-terminal Tyrosine or a C-terminal Phenylalanine amino acid;
  - (k) a *Streptococcus pneumoniae* polypeptide having a tripeptide RGD amino acid sequence;
  - (l) a *Streptococcus pneumoniae* polypeptide identified by proteomics as surface exposed;
- and
- (m) a *Streptococcus pneumoniae* polypeptide identified by proteomics as membrane associated.
46. The polypeptide of claim 45, wherein the polypeptide having a 0, 1 or 2 transmembrane domains comprises an amino acid sequence chosen from one of SEQ ID NO: 216, SEQ ID NO: 218, SEQ ID NO: 219, SEQ ID NO: 222, SEQ ID NO: 223, SEQ ID NO: 224, SEQ ID NO: 226, SEQ ID NO: 228, SEQ ID NO: 231, SEQ ID NO: 232, SEQ ID NO: 233, SEQ ID NO: 234, SEQ ID NO: 237, SEQ ID NO: 238, SEQ ID NO: 239, SEQ ID NO: 240, SEQ ID NO: 243, SEQ ID NO: 244, SEQ ID NO: 247, SEQ ID NO: 249, SEQ ID NO:

251, SEQ ID NO: 254, SEQ ID NO: 256, SEQ ID NO: 257, SEQ ID NO: 260, SEQ ID NO: 262, SEQ ID NO: 264, SEQ ID NO: 265, SEQ ID NO: 266, SEQ ID NO: 268, SEQ ID NO: 270, SEQ ID NO: 272, SEQ ID NO: 273, SEQ ID NO: 275, SEQ ID NO: 276, SEQ ID NO: 277, SEQ ID NO: 278, SEQ ID NO: 279, SEQ ID NO: 281, SEQ ID NO: 282, SEQ ID NO: 283, SEQ ID NO: 284, SEQ ID NO: 285, SEQ ID NO: 286, SEQ ID NO: 287, SEQ ID NO: 289, SEQ ID NO: 293, SEQ ID NO: 294, SEQ ID NO: 296, SEQ ID NO: 298, SEQ ID NO: 300, SEQ ID NO: 301, SEQ ID NO: 304, SEQ ID NO: 306, SEQ ID NO: 307, SEQ ID NO: 310, SEQ ID NO: 311, SEQ ID NO: 312, SEQ ID NO: 315, SEQ ID NO: 319, SEQ ID NO: 320, SEQ ID NO: 321, SEQ ID NO: 324, SEQ ID NO: 325, SEQ ID NO: 326, SEQ ID NO: 328, SEQ ID NO: 331, SEQ ID NO: 336, SEQ ID NO: 337, SEQ ID NO: 338, SEQ ID NO: 340, SEQ ID NO: 341, SEQ ID NO: 342, SEQ ID NO: 343, SEQ ID NO: 346, SEQ ID NO: 347, SEQ ID NO: 349, SEQ ID NO: 351, SEQ ID NO: 352, SEQ ID NO: 353, SEQ ID NO: 356, SEQ ID NO: 357, SEQ ID NO: 358, SEQ ID NO: 359, SEQ ID NO: 362, SEQ ID NO: 363, SEQ ID NO: 364, SEQ ID NO: 365, SEQ ID NO: 370, SEQ ID NO: 371, SEQ ID NO: 373, SEQ ID NO: 376, SEQ ID NO: 377, SEQ ID NO: 380, SEQ ID NO: 385, SEQ ID NO: 386, SEQ ID NO: 387, SEQ ID NO: 389, SEQ ID NO: 391, SEQ ID NO: 394, SEQ ID NO: 398, SEQ ID NO: 400, SEQ ID NO: 402, SEQ ID NO: 407, SEQ ID NO: 410, SEQ ID NO: 411, SEQ ID NO: 412, SEQ ID NO: 414, SEQ ID NO: 415, SEQ ID NO: 416, SEQ ID NO: 417, SEQ ID NO: 419, SEQ ID NO: 420, SEQ ID NO: 422, SEQ ID NO: 424, SEQ ID NO: 425], a biological equivalent thereof, or a fragment thereof.

47. The polypeptide of claim 45, wherein the polypeptide having 3 or more transmembrane domains comprises an amino acid sequence chosen from one of SEQ ID NO: 217, SEQ ID NO: 220, SEQ ID NO: 221, SEQ ID NO: 225, SEQ ID NO: 227, SEQ ID NO: 229, SEQ ID NO: 230, SEQ ID NO: 235, SEQ ID NO: 236, SEQ ID NO: 241, SEQ ID NO: 242, SEQ ID NO: 245, SEQ ID NO: 246, SEQ ID NO: 248, SEQ ID NO: 250, SEQ ID NO: 252, SEQ ID NO: 253, SEQ ID NO: 255, SEQ ID NO: 258, SEQ ID NO: 259, SEQ ID NO: 261, SEQ ID NO: 263, SEQ ID NO: 267, SEQ ID NO: 269, SEQ ID NO: 271,

SEQ ID NO: 274, SEQ ID NO: 280, SEQ ID NO: 286, SEQ ID NO: 290, SEQ ID NO: 291, SEQ ID NO: 292, SEQ ID NO: 295, SEQ ID NO: 297, SEQ ID NO: 299, SEQ ID NO: 302, SEQ ID NO: 303, SEQ ID NO: 305, SEQ ID NO: 308, SEQ ID NO: 309, SEQ ID NO: 313, SEQ ID NO: 314, SEQ ID NO: 316, SEQ ID NO: 317, SEQ ID NO: 318, SEQ ID NO: 322, SEQ ID NO: 323, SEQ ID NO: 327, SEQ ID NO: 329, SEQ ID NO: 330, SEQ ID NO: 332, SEQ ID NO: 333, SEQ ID NO: 334, SEQ ID NO: 335, SEQ ID NO: 339, SEQ ID NO: 344, SEQ ID NO: 345, SEQ ID NO: 348, SEQ ID NO: 350, SEQ ID NO: 354, SEQ ID NO: 355, SEQ ID NO: 360, SEQ ID NO: 361, SEQ ID NO: 366, SEQ ID NO: 367, SEQ ID NO: 368, SEQ ID NO: 369, SEQ ID NO: 372, SEQ ID NO: 374, SEQ ID NO: 375, SEQ ID NO: 378, SEQ ID NO: 379, SEQ ID NO: 381, SEQ ID NO: 382, SEQ ID NO: 383, SEQ ID NO: 384, SEQ ID NO: 388, SEQ ID NO: 390, SEQ ID NO: 392, SEQ ID NO: 393, SEQ ID NO: 395, SEQ ID NO: 396, SEQ ID NO: 397, SEQ ID NO: 399, SEQ ID NO: 401, SEQ ID NO: 403, SEQ ID NO: 404, SEQ ID NO: 405, SEQ ID NO: 406, SEQ ID NO: 408, SEQ ID NO: 409, SEQ ID NO: 413, SEQ ID NO: 418, SEQ ID NO: 421, SEQ ID NO: 423, SEQ ID NO: 426, SEQ ID NO: 427, SEQ ID NO: 428, SEQ ID NO: 429, SEQ ID NO: 430, a biological equivalent thereof, or a fragment thereof.

48. The polypeptide of claim 45, wherein the polypeptide having an outer membrane domain or a periplasmic domain comprises an amino acid sequence chosen from one of SEQ ID NO: 218, SEQ ID NO: 223, SEQ ID NO: 224, SEQ ID NO: 238, SEQ ID NO: 254, SEQ ID NO: 265, SEQ ID NO: 277, SEQ ID NO: 282, SEQ ID NO: 293, SEQ ID NO: 300, SEQ ID NO: 340, SEQ ID NO: 349, SEQ ID NO: 362, SEQ ID NO: 380, SEQ ID NO: 387, SEQ ID NO: 394, a biological equivalent thereof, or a fragment thereof.
49. The polypeptide of claim 45, wherein the polypeptide having an inner membrane domain comprises an amino acid sequence chosen from one of SEQ ID NO: 217, SEQ ID NO: 220, SEQ ID NO: 221, SEQ ID NO: 222, SEQ ID NO: 225, SEQ ID NO: 226, SEQ ID NO: 227, SEQ ID NO: 228, SEQ ID NO: 229, SEQ ID NO: 230, SEQ ID NO: 231, SEQ ID NO: 232, SEQ ID NO:

234, SEQ ID NO: 235, SEQ ID NO: 236, SEQ ID NO: 237, SEQ ID NO: 241, SEQ ID NO: 242, SEQ ID NO: 243, SEQ ID NO: 244, SEQ ID NO: 245, SEQ ID NO: 246, SEQ ID NO: 247, SEQ ID NO: 248, SEQ ID NO: 249, SEQ ID NO: 250, SEQ ID NO: 251, SEQ ID NO: 252, SEQ ID NO: 253, SEQ ID NO: 255, SEQ ID NO: 258, SEQ ID NO: 259, SEQ ID NO: 261, SEQ ID NO: 262, SEQ ID NO: 263, SEQ ID NO: 266, SEQ ID NO: 267, SEQ ID NO: 268, SEQ ID NO: 269, SEQ ID NO: 271, SEQ ID NO: 274, SEQ ID NO: 275, SEQ ID NO: 276, SEQ ID NO: 280, SEQ ID NO: 283, SEQ ID NO: 284, SEQ ID NO: 285, SEQ ID NO: 286, SEQ ID NO: 288, SEQ ID NO: 290, SEQ ID NO: 291, SEQ ID NO: 292, SEQ ID NO: 294, SEQ ID NO: 295, SEQ ID NO: 296, SEQ ID NO: 297, SEQ ID NO: 298, SEQ ID NO: 299, SEQ ID NO: 301, SEQ ID NO: 302, SEQ ID NO: 303, SEQ ID NO: 305, SEQ ID NO: 306, SEQ ID NO: 308, SEQ ID NO: 309, SEQ ID NO: 310, SEQ ID NO: 311, SEQ ID NO: 312, SEQ ID NO: 313, SEQ ID NO: 314, SEQ ID NO: 315, SEQ ID NO: 316, SEQ ID NO: 317, SEQ ID NO: 318, SEQ ID NO: 320, SEQ ID NO: 321, SEQ ID NO: 322, SEQ ID NO: 323, SEQ ID NO: 324, SEQ ID NO: 327, SEQ ID NO: 328, SEQ ID NO: 329, SEQ ID NO: 330, SEQ ID NO: 332, SEQ ID NO: 333, SEQ ID NO: 334, SEQ ID NO: 335, SEQ ID NO: 336, SEQ ID NO: 337, SEQ ID NO: 338, SEQ ID NO: 339, SEQ ID NO: 341, SEQ ID NO: 342, SEQ ID NO: 343, SEQ ID NO: 344, SEQ ID NO: 345, SEQ ID NO: 346, SEQ ID NO: 347, SEQ ID NO: 348, SEQ ID NO: 350, SEQ ID NO: 351, SEQ ID NO: 354, SEQ ID NO: 355, SEQ ID NO: 356, SEQ ID NO: 357, SEQ ID NO: 359, SEQ ID NO: 360, SEQ ID NO: 361, SEQ ID NO: 362, SEQ ID NO: 365, SEQ ID NO: 366, SEQ ID NO: 367, SEQ ID NO: 368, SEQ ID NO: 369, SEQ ID NO: 371, SEQ ID NO: 372, SEQ ID NO: 373, SEQ ID NO: 374, SEQ ID NO: 375, SEQ ID NO: 377, SEQ ID NO: 378, SEQ ID NO: 379, SEQ ID NO: 381, SEQ ID NO: 382, SEQ ID NO: 383, SEQ ID NO: 384, SEQ ID NO: 385, SEQ ID NO: 388, SEQ ID NO: 390, SEQ ID NO: 391, SEQ ID NO: 392, SEQ ID NO: 393, SEQ ID NO: 395, SEQ ID NO: 396, SEQ ID NO: 397, SEQ ID NO: 399, SEQ ID NO: 401, SEQ ID NO: 402, SEQ ID NO: 403, SEQ ID NO: 404, SEQ ID NO: 405, SEQ ID NO: 406, SEQ ID NO: 407, SEQ ID NO: 408, SEQ ID NO: 409, SEQ ID NO: 410, SEQ ID NO: 413, SEQ ID NO: 415, SEQ ID NO: 418, SEQ ID NO: 421, SEQ ID NO: 423, SEQ ID NO: 424, SEQ ID NO: 426,

SEQ ID NO: 427, SEQ ID NO: 428, SEQ ID NO: 429, SEQ ID NO: 430, a biological equivalent thereof, or a fragment thereof.

50. The polypeptide of claim 45, wherein the polypeptide identified by Blastp analysis comprises an amino acid sequence chosen from one of SEQ ID NO: 216, SEQ ID NO: 217, SEQ ID NO: 222, SEQ ID NO: 225, SEQ ID NO: 227, SEQ ID NO: 231, SEQ ID NO: 235, SEQ ID NO: 239, SEQ ID NO: 242, SEQ ID NO: 245, SEQ ID NO: 246, SEQ ID NO: 247, SEQ ID NO: 248, SEQ ID NO: 249, SEQ ID NO: 250, SEQ ID NO: 253, SEQ ID NO: 255, SEQ ID NO: 257, SEQ ID NO: 258, SEQ ID NO: 259, SEQ ID NO: 263, SEQ ID NO: 266, SEQ ID NO: 268, SEQ ID NO: 269, SEQ ID NO: 275, SEQ ID NO: 276, SEQ ID NO: 280, SEQ ID NO: 282, SEQ ID NO: 283, SEQ ID NO: 284, SEQ ID NO: 285, SEQ ID NO: 286, SEQ ID NO: 290, SEQ ID NO: 291, SEQ ID NO: 292, SEQ ID NO: 293, SEQ ID NO: 294, SEQ ID NO: 295, SEQ ID NO: 302, SEQ ID NO: 303, SEQ ID NO: 305, SEQ ID NO: 309, SEQ ID NO: 310, SEQ ID NO: 311, SEQ ID NO: 313, SEQ ID NO: 315, SEQ ID NO: 318, SEQ ID NO: 320, SEQ ID NO: 322, SEQ ID NO: 323, SEQ ID NO: 324, SEQ ID NO: 327, SEQ ID NO: 328, SEQ ID NO: 330, SEQ ID NO: 332, SEQ ID NO: 333, SEQ ID NO: 337, SEQ ID NO: 338, SEQ ID NO: 339, SEQ ID NO: 342, SEQ ID NO: 344, SEQ ID NO: 346, SEQ ID NO: 347, SEQ ID NO: 348, SEQ ID NO: 349, SEQ ID NO: 350, SEQ ID NO: 351, SEQ ID NO: 353, SEQ ID NO: 354, SEQ ID NO: 356, SEQ ID NO: 359, SEQ ID NO: 361, SEQ ID NO: 362, SEQ ID NO: 366, SEQ ID NO: 367, SEQ ID NO: 369, SEQ ID NO: 370, SEQ ID NO: 372, SEQ ID NO: 373, SEQ ID NO: 374, SEQ ID NO: 375, SEQ ID NO: 376, SEQ ID NO: 377, SEQ ID NO: 378, SEQ ID NO: 380, SEQ ID NO: 381, SEQ ID NO: 382, SEQ ID NO: 384, SEQ ID NO: 387, SEQ ID NO: 388, SEQ ID NO: 391, SEQ ID NO: 392, SEQ ID NO: 393, SEQ ID NO: 395, SEQ ID NO: 396, SEQ ID NO: 397, SEQ ID NO: 399, SEQ ID NO: 400, SEQ ID NO: 401, SEQ ID NO: 403, SEQ ID NO: 404, SEQ ID NO: 406, SEQ ID NO: 408, SEQ ID NO: 411, SEQ ID NO: 412, SEQ ID NO: 413, SEQ ID NO: 414, SEQ ID NO: 415, SEQ ID NO: 416, SEQ ID NO: 417, SEQ ID NO: 419, SEQ ID NO: 420, SEQ ID NO: 421, SEQ ID NO: 422, SEQ ID NO: 423, SEQ ID



NO: 425, SEQ ID NO: 427, SEQ ID NO: 428, SEQ ID NO: 429, a biological equivalent thereof, or a fragment thereof.

51. The polypeptide of claim 45, wherein the polypeptide identified by Pfam analysis comprises an amino acid sequence chosen from one of SEQ ID NO: 219, SEQ ID NO: 233, SEQ ID NO: 234, SEQ ID NO: 255, SEQ ID NO: 260, SEQ ID NO: 270, SEQ ID NO: 272, SEQ ID NO: 273, SEQ ID NO: 278, SEQ ID NO: 279, SEQ ID NO: 281, SEQ ID NO: 287, SEQ ID NO: 289, SEQ ID NO: 304, SEQ ID NO: 307, SEQ ID NO: 319, SEQ ID NO: 326, SEQ ID NO: 331, SEQ ID NO: 334, SEQ ID NO: 343, SEQ ID NO: 352, SEQ ID NO: 357, SEQ ID NO: 358, SEQ ID NO: 364, SEQ ID NO: 366, SEQ ID NO: 367, SEQ ID NO: 368, SEQ ID NO: 372, SEQ ID NO: 374, SEQ ID NO: 375, SEQ ID NO: 377, SEQ ID NO: 378, SEQ ID NO: 379, SEQ ID NO: 380, SEQ ID NO: 381, SEQ ID NO: 384, SEQ ID NO: 386, SEQ ID NO: 389, SEQ ID NO: 391, SEQ ID NO: 395, SEQ ID NO: 397, SEQ ID NO: 398, SEQ ID NO: 399, SEQ ID NO: 401, SEQ ID NO: 403, SEQ ID NO: 404, SEQ ID NO: 410, SEQ ID NO: 413, SEQ ID NO: 414, SEQ ID NO: 420, SEQ ID NO: 427, SEQ ID NO: 428, a biological equivalent thereof, or a fragment thereof.
52. The polypeptide of claim 45, wherein the polypeptide is a lipoprotein, the polypeptide comprises an amino acid sequence chosen from one of SEQ ID NO: 218, SEQ ID NO: 223, SEQ ID NO: 224, SEQ ID NO: 228, SEQ ID NO: 236, SEQ ID NO: 241, SEQ ID NO: 249, SEQ ID NO: 277, SEQ ID NO: 282, SEQ ID NO: 300, SEQ ID NO: 349, SEQ ID NO: 362, SEQ ID NO: 365, SEQ ID NO: 383, SEQ ID NO: 385, SEQ ID NO: 388, a biological equivalent thereof, or a fragment thereof.
53. The polypeptide of claim 45, wherein the polypeptide having a LPXTG motif and covalently associated with the peptidoglycan layer comprises an amino acid sequence chosen from one of SEQ ID NO: 228, SEQ ID NO: 236, SEQ ID NO: 249, SEQ ID NO: 385, a biological equivalent thereof, or a fragment thereof.

54. The polypeptide of claim 45, wherein the polypeptide having a peptidoglycan binding motif and associated with the peptidoglycan layer comprises an amino acid sequence chosen from one of SEQ ID NO: 240, SEQ ID NO: 264, SEQ ID NO: 325, a biological equivalent thereof, or a fragment thereof.
55. The polypeptide of claim 45, wherein the polypeptide having a signal sequence and a C-terminal Tyrosine or Phenylalanine amino acid comprises an amino acid sequence chosen from one of SEQ ID NO:226, SEQ ID NO:254, SEQ ID NO:289, SEQ ID NO:312, SEQ ID NO:321, SEQ ID NO: 340, SEQ ID NO:402, a biological equivalent thereof, or a fragment thereof.
56. The polypeptide of claim 45, wherein the polypeptide having a tripeptide RGD sequence that potentially is involved in cell attachment comprises an amino acid sequence chosen from one of SEQ ID NO:216, SEQ ID NO:236, SEQ ID NO:281, SEQ ID NO:282, a biological equivalent thereof, or a fragment thereof.
57. The polypeptide of claim 45, wherein the polypeptide identified by proteomics as surface exposed comprises an amino acid sequence chosen from one of SEQ ID NO: 229, SEQ ID NO: 231, SEQ ID NO: 232, SEQ ID NO: 261, SEQ ID NO: 279, SEQ ID NO: 281, SEQ ID NO: 282, SEQ ID NO: 284, SEQ ID NO: 286, SEQ ID NO: 289, SEQ ID NO: 306, SEQ ID NO: 318, SEQ ID NO: 331, SEQ ID NO: 343, SEQ ID NO: 346, SEQ ID NO: 351, SEQ ID NO: 366, SEQ ID NO: 371, SEQ ID NO: 374, SEQ ID NO: 377, SEQ ID NO: 379, SEQ ID NO: 387, SEQ ID NO: 391, SEQ ID NO: 393, SEQ ID NO: 394, 395, SEQ ID NO: 397, SEQ ID NO: 420, a biological equivalent thereof, or a fragment thereof.
58. The polypeptide of claim 45, wherein the polypeptide identified by proteomics as surface exposed comprises an amino acid sequence chosen from one of SEQ ID NO:592 through SEQ ID NO:752, a biological equivalent thereof, or a fragment thereof.

59. A recombinant expression vector comprising a nucleotide sequence having at least about 95% identity to a nucleotide sequence chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO: 431 through SEQ ID NO: 591, a degenerate variant thereof, or a fragment thereof.
60. The vector of claim 59, wherein the polynucleotide is selected from the group consisting of DNA, chromosomal DNA, cDNA, RNA and antisense RNA.
61. The vector of claim 60, wherein the polynucleotide comprises heterologous nucleotide sequences.
62. The vector of claim 61, wherein the polynucleotide is operatively linked to one or more gene expression regulatory elements.
63. The vector of claim 62, wherein the polynucleotide encodes a polypeptide comprising an amino acid sequence having at least about 95% identity to an amino acid sequence chosen from one of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID NO: 592 through SEQ ID NO: 752, a biological equivalent thereof, or a fragment thereof.
64. The vector of claim 59, wherein the vector is a plasmid.
65. A genetically engineered host cell, transfected, transformed or infected with the vector of claim 59.
66. The host cell of claim 65, wherein the host cell is a bacterial cell.
67. The host cell of claim 66, wherein the polynucleotide is expressed to produce the encoded polypeptide, a biological equivalent thereof, or a fragment thereof.
68. An antibody specific for a *Streptococcus pneumoniae* polynucleotide chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO: 431

through SEQ ID NO: 591, a fragment thereof, a degenerate variant thereof, or a *Streptococcus pneumoniae* polypeptide chosen from one of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID NO: 592 through SEQ ID NO: 752, a biological equivalent thereof, or a fragment thereof.

69. The antibody of claim 68, wherein the antibody is selected from the group consisting of monoclonal, polyclonal, chimeric, humanized and single chain.
70. The antibody of claim 69, wherein the antibody is monoclonal.
71. The antibody of claim 70, wherein the antibody is humanized.
72. An immunogenic composition comprising a polypeptide having an amino acid sequence chosen from one or more of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID NO: 592 through SEQ ID NO: 752, a biological equivalent thereof, or a fragment thereof.
73. The immunogenic composition of claim 72, further comprising a pharmaceutically acceptable carrier.
74. The immunogenic composition of claim 72, further comprising one or more adjuvants.
75. The immunogenic composition of claim 72, wherein the polypeptide is further defined as:
  - (a) a *Streptococcus pneumoniae* polypeptide having 0, 1 or 2 transmembrane domains;
  - (b) a *Streptococcus pneumoniae* polypeptide having 3 or more transmembrane domains;
  - (c) a *Streptococcus pneumoniae* polypeptide having an outer membrane domain or a periplasmic domain;
  - (d) a *Streptococcus pneumoniae* polypeptide having an inner membrane domain;

- (e) a *Streptococcus pneumoniae* polypeptide identified by Blastp analysis;
  - (f) a *Streptococcus pneumoniae* polypeptide identified by Pfam analysis;
  - (g) a *Streptococcus pneumoniae* lipoprotein;
  - (h) a *Streptococcus pneumoniae* polypeptide having a LPXTG motif, wherein the polypeptide is covalently attached to the peptidoglycan layer;
  - (i) a *Streptococcus pneumoniae* polypeptide having a peptidoglycan binding motif, wherein the polypeptide is associated with the peptidoglycan layer;
  - (j) a *Streptococcus pneumoniae* polypeptide having a signal sequence and a C-terminal Tyrosine or a C-terminal Phenylalanine amino acid;
  - (k) a *Streptococcus pneumoniae* polypeptide having a tripeptide RGD amino acid sequence;
  - (l) a *Streptococcus pneumoniae* polypeptide identified by proteomics as surface exposed;
- and
- (m) a *Streptococcus pneumoniae* polypeptide identified by proteomics as membrane associated.
76. The immunogenic composition of claim 75, wherein the polypeptide further comprises heterologous amino acids.
77. The immunogenic composition of claim 75, wherein the polypeptide is a fusion polypeptide.
78. The immunogenic composition of claim 75, wherein the polypeptide is encoded by a polynucleotide comprising a nucleotide sequence having at least about 95% identity to a nucleotide sequence chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO: 431 through SEQ ID NO: 591, a degenerate variant thereof, or a fragment thereof.
79. The immunogenic composition of claim 78, wherein the polynucleotide further comprises heterologous nucleotides.

80. An immunogenic composition comprising a polynucleotide having a nucleotide sequence chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO: 431 through SEQ ID NO: 591, a degenerate variant thereof, or a fragment thereof and is comprised in an expression vector.
81. The immunogenic composition of claim 80, wherein the vector is plasmid DNA.
82. The immunogenic composition of claim 81, wherein the polynucleotide comprises heterologous nucleotides.
83. The immunogenic composition of claim 82, wherein the polynucleotide is operatively linked to one or more gene expression regulatory elements.
84. The immunogenic composition of claim 83, wherein the polynucleotide directs the expression of a neutralizing epitope of *Streptococcus pneumoniae*.
85. The immunogenic composition of claim 84, further comprising one or more adjuvants.
86. A pharmaceutical composition comprising a polypeptide and a pharmaceutically acceptable carrier, wherein the polypeptide comprises an amino acid chosen from one of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID NO: 592 through SEQ ID NO: 752, a biological equivalent thereof, or a fragment thereof.
87. The pharmaceutical composition of claim 86, wherein the polypeptide is further defined as:
- (a) a *Streptococcus pneumoniae* polypeptide having 0, 1 or 2 transmembrane domains;
  - (b) a *Streptococcus pneumoniae* polypeptide having 3 or more transmembrane domains;

- (c) a *Streptococcus pneumoniae* polypeptide having an outer membrane domain or a periplasmic domain;
  - (d) a *Streptococcus pneumoniae* polypeptide having an inner membrane domain;
  - (e) a *Streptococcus pneumoniae* polypeptide identified by Blastp analysis;
  - (f) a *Streptococcus pneumoniae* polypeptide identified by Pfam analysis;
  - (g) a *Streptococcus pneumoniae* lipoprotein;
  - (h) a *Streptococcus pneumoniae* polypeptide having a LPXTG motif, wherein the polypeptide is covalently attached to the peptidoglycan layer;
  - (i) a *Streptococcus pneumoniae* polypeptide having a peptidoglycan binding motif, wherein the polypeptide is associated with the peptidoglycan layer;
  - (j) a *Streptococcus pneumoniae* polypeptide having a signal sequence and a C-terminal Tyrosine or a C-terminal Phenylalanine amino acid;
  - (k) a *Streptococcus pneumoniae* polypeptide having a tripeptide RGD amino acid sequence;
  - (l) a *Streptococcus pneumoniae* polypeptide identified by proteomics as surface exposed;
- and
- (m) a *Streptococcus pneumoniae* polypeptide identified by proteomics as membrane associated.
88. The pharmaceutical composition of claim 87, wherein the polypeptide further comprises heterologous amino acids.
89. The pharmaceutical composition of claim 87, wherein the polypeptide is a fusion polypeptide.
90. A DNA chip comprising an array of polynucleotides, wherein at least one of the polynucleotides comprise a nucleotide sequence chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO: 431 through SEQ ID NO: 591, a complement thereof, a degenerate variant thereof, or a fragment thereof.

91. A protein chip comprising an array of polypeptides, wherein at least one of the polypeptides comprises an amino acid sequence chosen from one of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID NO: 592 through SEQ ID NO: 752, a biological equivalent thereof, or a fragment thereof.
92. A method of immunizing against *Streptococcus pneumoniae* comprising administering to a host an immunizing amount of an immunogenic composition comprising a polypeptide and a pharmaceutically acceptable carrier, wherein the polypeptide comprises an amino acid sequence chosen from one or more of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID NO: 592 through SEQ ID NO: 752, a biological equivalent thereof, or a fragment thereof.
93. The method of claim 92, wherein the polypeptide is a fusion polypeptide.
94. The method of claim 92, further comprising an adjuvant.
95. A method for the detection and/or identification of *Streptococcus pneumoniae* in a biological sample comprising:
- (a) contacting the sample with an oligonucleotide probe of a polynucleotide comprising the nucleotide sequence chosen from one of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO: 431 through SEQ ID NO: 591, a degenerate variant thereof, or a fragment thereof, under conditions permitting hybridization; and
  - (b) detecting the presence of hybridization complexes in the sample, wherein hybridization complexes indicate the presence of *Streptococcus pneumoniae* in the sample.
96. A method for the detection and/or identification of *Streptococcus pneumoniae* in a biological sample comprising:
- (a) contacting the sample with an oligonucleotide primer of a polynucleotide comprising the nucleotide sequence chosen from one



- of SEQ ID NO: 1 through SEQ ID NO: 215 or SEQ ID NO: 431 through SEQ ID NO: 591, a degenerate variant thereof, or a fragment thereof, in the presence of nucleotides and a polymerase enzyme under conditions permitting primer extension; and
- (b) detecting the presence of primer extension products in the sample, wherein extension products indicate the presence of *Streptococcus pneumoniae* in the sample.
97. A method for the detection and/or identification of *Streptococcus pneumoniae* in a biological sample comprising:
- (a) contacting the sample with an antibody specific for a polypeptide comprising an amino acid sequence chosen from one of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID NO: 592 through SEQ ID NO: 752, a biological equivalent thereof, or a fragment thereof, under conditions permitting immune complex formation; and
- (b) detecting the presence of immune complexes in the sample, wherein immune complexes indicate the presence of *Streptococcus pneumoniae* in the sample.
98. A method for the detection and/or identification of antibodies to *Streptococcus pneumoniae* in a biological sample comprising:
- (a) contacting the sample with a polypeptide comprising an amino acid sequence chosen from one of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID NO: 592 through SEQ ID NO: 752, a biological equivalent thereof, or a fragment thereof, under conditions permitting immune complex formation; and
- (b) detecting the presence of immune complexes in the sample, wherein immune complexes indicate the presence of *Streptococcus pneumoniae* in the sample.
99. A kit comprising a container containing an isolated polynucleotide comprising a nucleotide sequence chosen from one of SEQ ID NO: 1 through SEQ ID

NO: 215 or SEQ ID NO: 431 through SEQ ID NO: 591, a degenerate variant thereof, or a fragment thereof.

100. The kit of claim 99, wherein the polynucleotide is a primer or a probe.
101. The kit of claim 100, wherein the polynucleotide is a primer and the kit further comprises a container containing a polymerase.
102. The kit of claim 99, wherein the kit further comprises a container containing dNTP.
103. A kit comprising a container containing an antibody that immunospecifically binds to a polypeptide comprising the amino acid sequence chosen from one of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID NO: 592 through SEQ ID NO: 752, a biological equivalent thereof, or a fragment thereof.
104. A kit comprising a container containing an antibody that immunospecifically binds to a fusion polypeptide comprising at least the amino acid sequence chosen from one of SEQ ID NO: 216 through SEQ ID NO: 430 or SEQ ID NO: 592 through SEQ ID NO: 752, a biological equivalent thereof, or a fragment thereof.
105. A method for producing a polypeptide which comprises culturing the genetically engineered host cell of claim 66 under conditions suitable to produce the polypeptide and recovering the polypeptide from the culture.

## SEQUENCE LISTING

&lt;110&gt; Wyeth

<120> NOVEL STREPTOCOCCUS PNEUMONIAE OPEN READING FRAMES ENCODING  
POLYPEPTIDE ANTIGENS AND USES THEREOF

&lt;130&gt; AM100649-PCT

&lt;160&gt; 752

&lt;170&gt; PatentIn version 3.1

&lt;210&gt; 1

&lt;211&gt; 684

&lt;212&gt; DNA

&lt;213&gt; Streptococcus pneumoniae

&lt;400&gt; 1

```

gctcgggcta aatcagtcga ctggactgat ttactacacc agtatagctt caagctctgt      60
cagaaacgat tctatcagcc cactgttcga atgcacttaa cccatcgagg agtacgagat      120
aagctgcttt ctactctgga gggattacag gttcactacg aactctatca actcctgctc      180
tttcattttc aagagaagaa tgccgaccat ttctttggat tgattgagca agaactgcca      240
acgggttcac cgctttttta aacggtcttt tggacttttt taagggatag agataagatt      300
atcaacgcac ttaagctgcc ttattccaac gctaaacttg aagcgaccaa taatttgatt      360
aagattatca agcgcaaaagc ctttggtttc cggaacttta acaattttta aaaacggatt      420
ttgatgactt tgaacatcaa aaaagagagt acgaatttcg tactctccag attgcagctt      480
ttcgctacc cactacactt gacaaagagc cactctttat tccatggtat caaaggcaag      540
acttggtttg gcatttaggt cccagcctgc gaagttttct ttgttccact cgctgacgct      600
ggcataggca atcatacctg cattgtctcc gcagagtcgc agagggggga tgataacctt      660
gacatctgtg atttcggctg ctag                                     684

```

&lt;210&gt; 2

&lt;211&gt; 675

&lt;212&gt; DNA

&lt;213&gt; Streptococcus pneumoniae

&lt;400&gt; 2

```

gagggggcgc aggcagccat gccaaaggct cttggctatg tcagtatcgg cctggcctgt      60
ggaattatcg gtgcgcctta tgtgacacct gttgagatgg gcttgatgag tctctttgtt      120

```

tatgctggga gtgccagtt tgccatgttg gcaactgattg tggttcaagc tccctgtggca	180
gctattgcta tgacgggttt tctaatacaac ttgcgtctct tttgttgag tttaacgca	240
tcgacttatt tccgtcatac cagtccttgg tacaatatcg gtatgtctag tatcttgaca	300
gatgagacct atggcggttt gatgggtgaa ttggcccata cagacaaggt aaatcctatg	360
tggatgcacg gaaacaatct taacagctat gtggcttggt ttgtggggac agtagtcgga	420
acggctcttg gtggcctgct accaaatcca gaaatctttg gcttggattt tgccttggt	480
gggatgttta ttggtatatt tgcctcgaa tttagatta tgcaaacg gattcctgtc	540
cgcaatctgc tcattatcct agcagttggt gcggtgtcct tcttttggct cttgacagtg	600
atgtctcagt cactagctgt tctgtttgcg acgctacttg gttgtagcat ggggttggt	660
ttagatgggc agtaa	675

&lt;210&gt; 3

&lt;211&gt; 864

&lt;212&gt; DNA

<213> *Streptococcus pneumoniae*

&lt;400&gt; 3

gattataagg tattctattt tggaggaaat gacatgaaa aaatcgtaa atactcatct	60
cttgagccc ttgctcttgt tgctgcaggt gtgcttgagg cttgctcagg ggggtgtaag	120
aaagaaggag aagcagctag caagaaagaa atcatcggtg caaccaatgg atcaccaaag	180
ccatttatct atgaagaaaa tggcgaattg actggttacg agattgaagt cgttcgcgct	240
atccttaag attctgacaa atatgatgtc aagtttgaaa agacagaatg gtcagggtgc	300
tttgctggtc ttgacgtga tcgttacaat atggctgtca acaatcttag ctacactaaa	360
gaacgtgcgg agaaatacct ctatgccga ccaattgcc aaaaatctaa tgccttgctc	420
gtgaagaaag atgactctag tatcaagtct ctcgatgata tcggtggaaa atcgacggaa	480
gtcgttaag ccactacatc agctaagcag ttagaagcat acaatgtga acacacggac	540
aacccaacta tccttaacta tactaaggca gacttgcaac aaatcatggt acggttgagc	600
gatggacaat ttgactataa gatttttgat aaaatcggtg ttgaacagat gatcaagaac	660
caaggttttg acaacttgaa agttatcgaa cttccaagc accaacaacc gtacgtttac	720
ccacttcttg ctcagggtca agatgagttg aaatcgtttg tagacaaacg catcaagaa	780
ctttataaag atggaactct tgaaaaattg tctaaacaat tcttcggaga cacttatcta	840

ccggcagaag ctgatattaa ataa 864

<210> 4  
 <211> 1389  
 <212> DNA  
 <213> *Streptococcus pneumoniae*

<400> 4  
 aaaggtagag agaatatggt ttttcctagc gaacaagaac agattgaaaa atttgaaaag 60  
 gatcatgtag ccagcatta ttttgagggt ttgcgtacct tgatttctaa gaaatcagtc 120  
 tttgccagc aggttggact caaggaagtc gcaaatatc tgggtgagat tttcaagcgt 180  
 gttggagctg aagtggagat tgatgagagc tatacagcgc cctttgtcat ggcacatttc 240  
 aagagtctgc gtccagatgc caagaccttg attttctata accactatga cactgtgcca 300  
 ggggatgggg atcaggtctg gacagagat ccttttacgc ttccggtccg caatggcttc 360  
 atgtatgggc gtggggttga tgacgacaag ggtcatatca cagctcgctt gagtgccttg 420  
 agaaaatata tgcagcacca tgatgattta cctgtcaata tcagctttat catggaggga 480  
 ggggaggaat cggcttcaac agacctagat aagtatttgg aaaagcatgc agacaaactc 540  
 cgtggggcgg atttgttggc ctgggaacaa gggaccaaaa atgccttgga acagctggaa 600  
 atttctggtg gcaataaggg gattgtgacc ttgatgcca aggtaaaaag cgctgatgtg 660  
 gatatccact cgagtatttg tgggtttgtg gaatcagctc cttggtatct cctccaagcc 720  
 ttacagctctc ttctgtctgc ggtgggcgt atcttgggtg aaggcttgta cgaagaagta 780  
 caagagccca atgaacgaga aatggccttg ctagaaactt atggccaacg aaaccagag 840  
 gaagttagtc ggtatttatg attggagtgt cctctcttac agggagagcg gatggccttt 900  
 ctaaaacgtt tctttttcga tccagcgctt aatatcgaag gaatccagtc tggttatcaa 960  
 ggtcaggggtg ttaagactat ttacctgca gaagccagt ccaagctaga ggttcgtctg 1020  
 gttccggggc tagaaccgca tgatgttctg gaaaaaatc ggaacagct agacaaaaat 1080  
 ggccttgata aggtagaatt atactatacc ttgggagaga tgagctatcg aagcgatatg 1140  
 agcgcaccag ccattctcaa tgtgatcgag ttggccaaga aattctatcc acagggcgctt 1200  
 tcagtcttgc cgacgacagc ggggacagga cctatgcata cggctcttga tgcctagag 1260  
 gtaccaatgg ttgcattcgg tctaggaaat gccaatagcc gagaccacgg tggagatgaa 1320  
 aatgtgcgaa tcgctgatta ttacaccoat atcgaattag tagaggagct gattagaagc 1380

tatgagtag 1389

<210> 5  
 <211> 624  
 <212> DNA  
 <213> *Streptococcus pneumoniae*

<400> 5  
 ggggaatatca tgggtagatt tttagacttt gtctttaatc gttctctttt agggatgatt 60  
 gcgacagcct tcttttggtt attaaacttta gcaggaggga ttatccttgg tctagcgccg 120  
 gctagtgccca ccttgatgag cttatatgca gaacatgggt atagctttcg ggaatacagt 180  
 ttgaaggagg cttggctctt ttacaagcaa aattttgtct caagcaacct gattttctat 240  
 agcttttttg gtgtgggtct agttttgacc tatggtttgt atctcttggg gcaattgcct 300  
 catcagacca ttgttcattt gattgcgacc cttttgaatg tcttagtagt tgcctgac 360  
 tttttggcct atacagtatc tttaaaatta caagtttatt ttgccttgc cttatcgaaat 420  
 agtctcaaat tatccttgat tggcatcttt atgagcttag cagctgtggc taaggttctc 480  
 cttgggactg tgctacttgt agcaattggt tattatatgc ctgccctgct attttttgta 540  
 ggaattggga tgtggcattt ctttatcagt gatagtgtgg aacctgtcta tgaatcctc 600  
 catgaaaaat tggcgacaaa atag 624

<210> 6  
 <211> 630  
 <212> DNA  
 <213> *Streptococcus pneumoniae*

<400> 6  
 actcttgcca accaatttta tccaaatttc ccaatcagaa atcatcaata tcgattccat 60  
 ctctcacctc aagctcacgc caaacggtct ggtagaaatt ttcttgaaa acgaaagctt 120  
 caccactctt tcacgcggtt atctaaaaac catcaaggag aaattagaac tatgaaaaaa 180  
 caagtatttc acgatgcagc taccggtggt cttatcggcc tctctctctc tctctcttt 240  
 tcaactcattt atgcacaaaa tacctacgca ccactaaatc cctactctct catagggcaa 300  
 gtgatggatc agcatcaggt tcacggtgcc ctggctcttc tctactgcac acttatctgg 360  
 gcaaccatcg gtatgctctt caactttggc aaccgcttat ttacccgtga ctggagcatg 420  
 cttcgtgccca ctctgactca tttcttctct atgctggctg gctttgtccc actagcaact 480  
 cttgctgggt ggttcccttt ccactggatt ttctacctcc agctcattat cgagtttggc 540

attgtctatc tcatactctg ggctattctc tataaaagag aggctaaaa agtagatcac 600  
atcaatcaac tcttgagca tagaaaatag 630

<210> 7  
<211> 609  
<212> DNA  
<213> Streptococcus pneumoniae

<400> 7  
gagagaagga atactatgta cgcataatga aaaggaatca ttacccaaat tactgcacaa 60  
tacattgttc ttgaaaccaa tggatattggg tataacctgc atgtggccaa tccttatgcc 120  
tattcaggtc aggttaatca ggaggctcag atttatgtgc atcaggttgt gcgtgaggac 180  
gcccatattgc tttatggatt tcgctcagag gatgagaaaa agctctttct tagtctgatt 240  
tcggtctctg ggatttggtc tgtatcagct cttgctatta tcgctgctga tgacaatgct 300  
ggcttggttc aagccattga aaccaagaac atcacctact tgaccaagtt ccctaaaatt 360  
ggcaagaaaa cagcccagca gatgggtgctg gacttggaag gcaaggtagt agttgcagga 420  
gatgaccttc ctgccagggt cgcagtgcaa gcaagtgctg aaaaccaaga attggaagaa 480  
gctatggaag ccatgttgcc tctgggctac aagccaacag agctcaagaa aatcaagaaa 540  
ttctttgaag gaacgacaga tacagctgag aactatatca agtcggccct taaaatgttg 600  
gtcaaatag 609

<210> 8  
<211> 675  
<212> DNA  
<213> Streptococcus pneumoniae

<400> 8  
tgtagaaat gcagaagcac gtttgcgtgc agctctataa acatcaaggc tgggagcact 60  
tccagctctt attctatttt aatttcaaaa agaagaaga aagaaatgaa aaaaatagtt 120  
cttgtagtc tagctttcct ttttgcctg gttgggtgcg gacagaaaaa agaactgga 180  
ccagctacaa aaacagaaaa agatacgtct cagtcggcat tgccagtatt tgaatatgct 240  
gagaagaata cagttgtaac taagactttg gtcttgcaca agtcagatga tggtagccag 300  
cagacacaaa caattactta caaagacaag acttttttga gtctagctat ccaacaaaaa 360  
cgccagctct ctgatgagtt gaagacttat attgaccaac atggagtgga ggaactcaa 420

aaagctcttc ttgaagcggg ggagaaggat aagtctatca ttgaagctcg taaattggca 480  
 ggtttcaaac ttgaacaac actattgagc gcaacggaac ttcaacaac gactagtttt 540  
 gattttcaag ttctggatgt caagaaggct tcccagttgg aacatctgaa gaatattggt 600  
 ttgaaaaatc ttttgaaaaa tgaaccaagc aaatatattt cagatagatt ggcaaatggc 660  
 gcgacagaac aatag 675

<210> 9  
 <211> 555  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 9  
 gcagataaat tgactccatt ttttgaactt gttatactag gggaattgct gggttagagaa 60  
 aatttctcta aattggtagc agaaaggaaa ttcatcatga aattaaaaag attcacactt 120  
 tctcttgctt ctctagcaag ttttagtctc ttagtagctt gttcacaag agctcaacag 180  
 gttcaacagc ctgttgctca gcagcaggtc caacaacctg ctcaacagaa taccaatact 240  
 gcaaatgcag gaggtaacca aatcaagcg gctccagtagc aaaaccaacc tggtgtctcaa 300  
 ccgaccgata ttgatgggac ttatactggt caggatgacg gagaccgtat cacttttagtg 360  
 gtaactggaa cgactggtac atggactgag ctgcaatctg acggggatca gaaagtcaaa 420  
 cagggtacat tggattcagc aatcaacgc atgattattg gcgatgatgt caaaatttac 480  
 actgtaaacg gtaatcaaat cgtcgtagat gatatggata gagaccatc ggaccaaatc 540  
 gttttaacta aataa 555

<210> 10  
 <211> 1557  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 10  
 cattcaaac atcaaggagg ggatatgaaa tataggaaat ttcaattatt gatgtccaag 60  
 tatggcttta gtctttcgat tatgctactt gaactttgtc ttgttttttg tctctttctt 120  
 tatttaggac gcattgctcc cattttatgg attactgtcc tcattctact gagtatcatc 180  
 acaatcattt cgatagtcac ccgtaatacg actcctgaga ataaggtaac ctggttgtta 240  
 gtagcctttg tgccagtatt tggcccttg ctctatctga tgtttggtga aaggcgattg 300  
 tccaaaaaag aatcaaaaa actgaagaag ctaggctcta tgcatttcca agaagcaaat 360



```

agccagctac taaaagagaa attaaaagaa agtgacaagg cagcttatgg agtcatcaag 420
tccttatatga gtatggatgac caatgctgac atctatgac aaactgcctc tacatttttt 480
cctaaccggag aagctatgtg gaaaagatg gtagaagatc ttaaaaaggc tgagaatttt 540
attttcttgg aatattacat tatagaagaa ggtttgatgt ggaatcgcat actagatata 600
ctagagcaaa aggtagctca ggggtgtagag gtaaatgac tctatgatga tatcgctgt 660
atggctactt taacaggaga ttatgcacat cgacttcgtc agctgggcat cgaggcccat 720
aaattcaata aagttattcc tcgtttgaca gtggcttata ataacagaga tcatagaaaa 780
atattgattg ttgatggtca gatagcctat actgggtggg tcaatctggc agatgagtac 840
attaaccacg tcgagagatt tggttattgg aaggatagtg gaattcgctt agacggacta 900
gcagtaaaag ctctgacag cttatttttg accacttggt acattaatcg aggagaaatt 960
agtgattttg atcaatatca tttagaaaat cattctatcc cgagtgcagg ttaaccatt 1020
ccatacggaa gtggaccaa gccaatttt cgagcgagg tagggaaaaa agtttatcag 1080
agttaatca atcaagcaac agaatcgggc tatattacga caccttattt gattatagat 1140
tatgatttaa cagagacaat caaaaatgca gctatgagag gggtcgatgt tcgaattatc 1200
accccttaca taccagataa gaagttcatt cagttagtc cagagaggag ttatcccgac 1260
cttctttctg ctggtgttgg gatttatgag tatagtcag gttttattca tagtaagcag 1320
atggttggtg acgaagattt tgcgggtggtg gggacaatca atctcgacta ccggagcttg 1380
gtacaccatt atgaaaatgc agtcttactc tataaaactc cttctataag ggaaatcgcc 1440
cgagattttc gaaatatatt tgcagattct caggaagtct atcctcatc tatcaaaacg 1500
agctggtatc aaaagcttgt aaaaagaaat gccagctat tcgccctat cttataa 1557

```

```

<210> 11
<211> 282
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 11
gaagacatca ttgatattt gattacctc gatgtcatga accaaacctt tggcacggta 60
gtgagcaatg attggttctc cttgagcaat attaacatcc aaacgacgtt ttactgtctc 120
aggcttatca tcttcacgtt ggtagtaatc ttctcttta tagtcaactg gtgggttaaa 180
gacctgtggg aaagtttctc cagttacgag gtggatgata cgccactca aacgttccaa 240

```

aaggctgtca gggttcactt caatattgat aacaccttct ag 282

<210> 12  
 <211> 1473  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 12  
 atgataaatg atataattct attattgttc gtaaaaatta aaaggagatt gatgatggac 60  
 aaattatttta aactaaaaga gaacggtaca gacgttcgta cagaggttct cgctggttta 120  
 acaactttct ttgcaatgag ctatattctc tttgtaaacc cacaaatact ttcacaaaca 180  
 ggaatgcctg ctacagggcgt ctctcctagcg acgattattg gtgcagtagc ggggtacctg 240  
 atgatggcct tttattgctaa ctacacctat gcccaagcgc caggtatggg actcaatgcc 300  
 ttctttacct ttacagttgt attcgggcctt gggtattctt ggcaagaagc cctagctatg 360  
 gtcttcatct gtgggattat ttcatgtatt attaccttga caaatgttcg taaatgatc 420  
 attgaatcga ttoccaatgc tcttcgtcga gctatttcag ctggtatcgg tgtcttctct 480  
 gcctatgtag ggataaagaa tgctggactt ttgaaattca cgattgatcc agggcaactat 540  
 actgtttagt gagaaggggc tgacaaagct caagcaacga ttgcagcaaa ctcttcagca 600  
 gttccaggat tggtcagcct taataatcca gctgttttag tggctcttgc aggacttgcc 660  
 attactatct tctttgtcat caaagggtt aaagggggaa ttattctctc tatcttgaca 720  
 acaactgttc ttgctattgc agttgggttg gttgatttgt ctagtatcga ttttgcataat 780  
 aacctagtgt gtgcagcctt tgaagatttg aagacaatct ttggtgcagc tcttggttca 840  
 gaaggatttg gagctttggt ttcatagata gctcgtctgc ctgaaactct gatggccatt 900  
 ctgccttctc cattgcacga tatctttgac acaattggta ccttgatcgg tacaggtgaa 960  
 aaagttggtg tcgtagcgac aaatggtgaa aatcaccaat cagccaaatt ggataaggct 1020  
 ctttactctg atttgatttg aacgacagtc ggtgccattg caggtacttc aaacgtaacg 1080  
 acttatgttg agtctctctc ttgtatcggt gcagggtggc gtactggttt gacagccttg 1140  
 gttgtagcta tctgttttgc gatttcaagc ttcttttagc cacttctagc gatcgtacca 1200  
 acagcggcta cagctccaat cttgattatc gttgggatta tgatgcttgg tagcttgaaa 1260  
 aatatccatt gggatgatat gtcgtaagca gttcctgcct tcttcacatc tatctttatg 1320  
 ggattcagct actctatcac tcaagggtt gcagttgggt tcttgactta cactttgact 1380

aagcttggtta aaggccaagt taaagatggt catgtcatga ttgggatttt ggcagccttg 1440  
 ttatccttâ actacatcag catggcctta taa 1473

<210> 13  
 <211> 3240  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 13  
 ttctattga ttttacaata tgtttattgg agtgtatata tgcaacaaa aacaaagaag 60  
 ctcatgtgta gtttgccttc acttgtttta tcaggatttt tattaacca ttatatgaca 120  
 attggagcgg aagaacgcac tacgaatacc attcagcaaa gccagaagga agttcagtat 180  
 cagcaaaagg atacaaaaaa tttagttgaa aatgggtgatt ttggtcagac ggaggacgga 240  
 agcagtcctg ggacaggaag caaagctcag ggggtggtcag ctggggtaga ccagaagaat 300  
 agtgcagatg cctcaactcg agtcattgag gctaaggatg gggctatcac tatctcaagc 360  
 catgagaaat taagggcagc gcttcaccgt atggttccta ttgaagctaa gaaaagtat 420  
 aaactgcgtt tcaagattaa aacagataat aaaatcgga ttgcaaagt tcgtatcatt 480  
 gaggaagt gtaaggacaa gcgatttggt aattctgcaa cgagtcagg aacaaaggac 540  
 tggcagacca ttgaagcaga ctatagccc actttagatg ttgataaaat caagctggag 600  
 ttattctatg aaacaggaa cgggactggt tcctttaagg atattgagct ggtagaggta 660  
 gcagaccagc tttctgagga ttctcaaca gataaacagc ttgaggaaaa gattgattta 720  
 ccaattggaa aaaaacatgt tttttctctt gcggactata ctataagggt agaaaatcct 780  
 gacgttgctt cagtcacaaa tggattttta gaacctctta aggaagggac aaccaatgtc 840  
 attgtcagta aagatggcaa ggaagtga aaagattcct tgaagattct ggcctctggt 900  
 aaggatgcat acacagaccg ttggatgac tggatggca tcacgctgg gaatcaatac 960  
 tatgattcta aaaatgaaca gatggccaaa ttaaaccagg aattggaagg aaaggtagct 1020  
 gatagcctat ccagtatctt aagtcaggcg gaccgcacct atttgggga aaaattttca 1080  
 aattataaaa cgtctgcaaa tctgactgcc acttatcgga aattggagga gatggccaag 1140  
 caagtgaaca atcctctctc tcgttattat caagatgaaa ctgtcgttcg aacagtcagg 1200  
 gattccatgg aatggatgca taaacatgtc tacaatagtg aaaagagcat tgttgggaac 1260  
 tgggtgggatt atgaatcgg tacacctcgt gccatcaaca atacctgtgc tctgatgaaa 1320

gaatacttct ctgatgagga aattaaaaaa tatacagatg tgattgaaaa atttgtacca	1380
gatcccgaaac atttcgaaa gacgactgat aaccattca aggcctcagg tggaaactta	1440
gttgatatgg gaagggtaaa agtaatatgt ggtttactgc gtaaggatga tcaagaaatt	1500
tcttctacca ttgcctgat tgagcaagtg ttcaagttgg tagaccaagg tgaaggtttt	1560
tatcaagatg gatcctatat cgaccacacc aatgttgctc atacgggtgc ttatgggaat	1620
gttttgattg atggcctgtc tcaactgttg ccagtcattc aaaagaccaa gaatccaatc	1680
gataaagata aaatgcaaac catgtaccac tggattgata aatcgtttgc tctttgtctg	1740
gtgaatggag agttgatgga tatgagtcgt ggacgctcga tcagtcgtgc aaatagcgag	1800
gggcacgtgg ccgcagtaga agtactaaga gggattcacc gaatagcgga tatgtctgaa	1860
ggagaaacca aacaatgttt gcagagtcct gtgaagacca ttgttcaatc ggatagtatt	1920
tatgatgtct ttaagaattt gaagacttat aaggatatca gtttgatgca atccttgta	1980
agtgatgcag gagtgcgaag tgttccaaga ccaagttacc tatctgcctt taacaagatg	2040
gataaaacag ccatgtacaa tgcagagaaa gggtttggat ttggcttctc actcttttcc	2100
agtcgtacct tgaattacga acacatgaac aaggaaaata aacgtggttg gtatacgagt	2160
gatgggatgt tctatcttta caatggcgat ttgagtcact atagcgatgg ctactggcca	2220
acagttaatc catataagat gcctggtaca acagagacgg atgctaagag agcggaatagc	2280
gatacaggta aagttttacc gtctgcttcc gttggaacga gcaactaga tgatgccaat	2340
gcgacgcaa ccatggattt cccaactgg aatcaaacat tgactgtcga taagagctgg	2400
tttatgctaa aggataagat cgccttttta ggaagcaata tccaaaacac ttcaacagat	2460
actgctgcaa ctacaattga ccagagaaaa ctggaatcag gtaatccata taaagtctat	2520
gtcaatgata aagaagcctc ccttacagaa caagaaaagg attatcctga aaccacaagt	2580
gtctttttag aatcgttcga ttcgaaaaag aatattggtt actttttctt taagaagagt	2640
tcaatcagta tgagttaagg tttgcaaaag ggagcctgga aggatatcaa tgaaggacag	2700
tcagacaagg aagttgaaaa tgaatttctt acgattatgc aggcctcataa gcaaatataga	2760
gattcttatg gctatatgct cattcctaac gtggatcgtg ccacottcaa tcaaatgata	2820
aaagagttag aaagtagcct catcgaaaat aacgaaaccc ttcagtcgtt ttatgatgct	2880
aaacaaggag tttggggcat tgtgaaatat gatgattctg tctctactat ttccaaccaa	2940

ttccaagttt tgaacgtgg agtctatacc attogaaaag aaggggatga atataagatt 3000  
gcctactata atctcgaaac ccaggaatca gctccagatc aggaagtctt taaaaagcta 3060  
gagcaagcag ctacgccaca agtacagaat tcaaaagaaa aggaaaaatc tgaagaggaa 3120  
aagaaccatt cggatcaaaa gaatctccct cagacaggag aaggtcagtc aatcttggca 3180  
agtctaggtt tcttgctact tggggcattt tatctattcc gtagaggaaa gaacaactaa 3240

<210> 14  
<211> 831  
<212> DNA  
<213> Streptococcus pneumoniae

<400> 14  
tggagctggt caagtcacaa ttatggacta tatttaaaag aggagatcgt tatgtcgatt 60  
aatgtatttc aagcgatttt aattggatta tggacagctt tctgttttag tggaaatgctg 120  
ttaggaattt acaccaatag atgtattgtt ctgtcatttg gtgtcggaaat tattctaggt 180  
gatctgccta ctgctcttgc aatgggagct attggtgaat tggcttatat gggattcgggt 240  
gttggtgctg gaggtactgt tccaccaaac ccaatcggaac ctggtatctt tggtaacctg 300  
atggctatca ctagtgtctg taaagtcagt ccagaagcgg ctcttgccct ctctactcgg 360  
attgctgtgg cgattcaatt cttacaaact ttccgcata ctgtacgtgc tgggtgcgcc 420  
gaaacagcta tgaagcactt gaaaaacat aatttgaaga aatttaagtt cactctaaat 480  
gcaacaattt ggttgtttgc ctttattgga ttaccttgg gttgcttggg tgccctttca 540  
atggatacct tgttgaaact cgtagactac attccaccgg tattacttac aggtttgaca 600  
gttgctggta aaatgctccc agctatcggg ttgtcgatga tctgttcagt gatggctaag 660  
aaagagtga ttccctttgt cttgttggga tatgtttgtg cagcttatct aaacatocca 720  
acaattggta ttgcaattgt aggtactatc ttgtcttga ttgaatttta taacaagcca 780  
aaaacagcgg atcatgtgtt agaggaggaa gcacacgatg actggatcta a 831

<210> 15  
<211> 399  
<212> DNA  
<213> Streptococcus pneumoniae

<400> 15  
tacatatgtt cgactcactt cgtattgcaa gagctaaaa agaccaggat taggaggtgc 60  
cttatgaaat cactactgtg actactgac attcatgttt ttatcagtat tttccttttc 120

ttcgccctta cttcaggagc tatttctcat acagttttac tactectact cctctttctt	180
cctgcgctca ataaaggact tgagaaaata caatcaaac ggatacctgt cctcaacgca	240
gcctcttctt ttctctcat atcttttcca caacttttaa ccaacctgt ccaatggaaa	300
ttttcaatat tctagtcgt aaccatcatt tcaagtttgg cctacttcta taacttttat	360
caagtagtta aagaagtaga tcaaaaacag ttgatttag	399

&lt;210&gt; 16

&lt;211&gt; 2256

&lt;212&gt; DNA

<213> *Streptococcus pneumoniae*

&lt;400&gt; 16

gatatgaagt ggacaaaaag agtaatcgt tatgcacca aaaatcgaa atcgccgct	60
gaaaacagac gcagagttag aaaaagctg agtttattat ctgtcttctt ttttgccatt	120
tttttagtca attttgcgt cattattggg acaggcactc gctttggaac agatttagcg	180
aaggaagcta agaaggttca tcaaacacc cgtacagtc ctgcaaacg tgggactatt	240
tatgaccgaa atggagtccc gattgctgag gatgcaacct cctataatgt ctatgcggtc	300
attgatgaga actataagtc agcaacgggt aagattcttt acgtagaaaa aacacaattt	360
aacaagggtg cagaggtctt tcataagtat ctggacatgg aagaatccta tgtaagagag	420
caactctcgc aacctaatct caagcaagtt tcttttgag caaagggaaa tgggattacc	480
tatgccata tgatgtctat caaaaaagaa ttggaagctg cagagggtcaa ggggattgat	540
tttacaacca gtcccaatcg tagttaccca aacggacaat ttgcttctag ttttatcggt	600
ctagctcagc tocatgaaaa tgaagatgga agcaagagct tgctgggaac ctctggaatg	660
gagagtctct tgaacagtat tcttgacagg acagacggca ttattacct tgaaaaaggat	720
cgctcgggta atattgtacc cggaacagaa caagtttccc aacgaacgat ggacggttag	780
gatgtttata caaccatttc cagccccctc cagtccttta tggaaaccca gatggatgct	840
tttcaagaga aggtaaaagg aaagtacatg acagcgactt tggtcagtgc taaaacaggg	900
gaaattcttg caacaacgca acgaccgacc tttgatgcag atacaaaaga aggcattaca	960
gaggactttg tttggcgtga tatcctttac caaagtaact atgagccagg ttccactatg	1020
aaagtgatga tgttgctgc tgctattgat aataatacct ttccaggagg agaagctctt	1080
aatagtagtg agttaaaaat tgcagatgcc acgattcgag attgggacgt taatgaagga	1140

```

ttgactgggtg gcagaacgat gactttttct caaggttttg cacactcaag taacgttggg 1200
atgacccctcc ttgagcaaaa gatgggagat gctacctggc ttgattatct taatcgtttt 1260
aaatttggag ttccgacccg tttcgggttg acggatgagt atgctgggtca gcttctctgcg 1320
gataatattg tcaacattgc gcaaagctca ttggacaag ggatttcagt gaccagacg 1380
caaatgattc gtgcctttac agctattgct aatgacgggt tcatgctgga gcctaaattt 1440
attagtgcga tttatgatcc aaatgatcaa actgctcgga aatctcaaaa agaaattgtg 1500
ggaaatcctg tttctaaga tgcagctagt ctaactcgga ctaacatggt ttgggtaggg 1560
acggatccgg tttatggaac catgtataac cacagcacag gcaagccaac tgtaactggt 1620
cctgggcaaa atgtagccct caagtctggt acggctcaga ttgctgacga gaaaaatggt 1680
ggttatctag tcgggttaac cgactatatt ttctcggtg tatcgatgag tccggctgaa 1740
aatcctgatt ttatcttgta tgtgacggtc caacaacctg aacattattc aggtattcag 1800
ttgggagaat ttgccaatcc tatcttgag cgggcttcag ctatgaaaga ctctctcaat 1860
cttcaacaaa cagctaaggc tttagagcaa gtaagtcaac aaagtcctta tcttatgcct 1920
agtgtcaagg atatttcacc tggtgattta gcagaagaat tgcgtcgcaa tcttgtaaaa 1980
cccatcgttg tgggaacagg aacgaagatt aaaaacagtt ctgctgaaga agggaagaat 2040
cttgcccgca accagcaagt ctttatctta tctgataaag cagaggagggt tccagatatg 2100
tatggttgga caaaggagac tggctgagacc cttgctaagt ggctcaatat agaacttgaa 2160
tttcaaggtt cgggctctac tgtgcagaag caagatgttc gtgctaacac agctatcaag 2220
gacattaaaa aaattacatt aactttagga gactaa 2256

```

<210> 17

<211> 660

<212> DNA

<213> *Streptococcus pneumoniae*

<400> 17

```

tttaatttgt caaatggaaa tagaatgaaa aatggaaa gaatttatag ttggagggtg 60
tttatgtacg gtaataaaa acgattagggt gatatatatt tatctttaat agggataata 120
atattgtgtc cggtttttat gataattgca attgcgatta aacttgattc agaagggtccg 180
gttatattta agcaaaaacg ctttgggtatt cataaagaat acttctatat ttgaaattt 240
aggctctatg aaatagatgc acctaaaaat gtggcgccctc gaaacttata taatccagag 300

```

caatggatta caaaagtagg ggctttcttg cgaaaaacat ctttggatga actaccacaa 360  
 ttgtttaata ttcttgttgg taatatgagt attgtaggtc ctagaccagc gggataataat 420  
 gaactagatt tgattgcaga gagagataag tatggagcaa atgatatctt gccagggtta 480  
 actggatggg cacaataaa cgggcgtgat accttgtctg ttgagatgaa gacggagtta 540  
 gatggctact atgttaaca tctgtctttg ataagtata tttagatgat agttaagaca 600  
 ataccttacg tactgaaacg aaaagggtatt gtagagggtta gtggaagaa agaaagttaa 660

<210> 18

<211> 1251

<212> DNA

<213> *Streptococcus pneumoniae*

<400> 18

gaaagaaagt taaattggac aatgaaaata ctatttgttt gccaacatta taagccagaa 60  
 ccattcaggt tgtcagatat ttgtgaagat ttagttcgaa aagggcata agtctctgtt 120  
 ttggctggga ttccctaatta cctgaaggg aagatatatg cagattatcg tcataataaa 180  
 aaaagacgtg agattataga aggtgttacg atatatcggt cttatacaat cctagaaaaa 240  
 aaaagtgttg talttcgatt gttgaattat tttagctttg caattagttc tacttttagga 300  
 gttttatttg ggaggataaa aacgaaagat ggatcgaatt ttgactgtgt attcgttaac 360  
 caattgtctc cagttatgat ggcattgggt ggtatggctt ataaaaaaa atataagaaa 420  
 ccgatgttcc tatattgtat ggatgtttgg ccagatagtt taaccgtagg tggagtgtgaa 480  
 caagatggct tgattttcaa gctgtttaaa tttatctcaa aaaagttta ccgagctagt 540  
 gattatatat ttgtcactag tccatcattt aaaaattatt ttgtgaagca atttgacata 600  
 tccgaacaaa agattacata ttggccacaa tatgcagaag atctttttat cctgatgaa 660  
 tctatagtta ataaagaaag tgttgacctt acctttgctg gtaatatgg caaagcacaa 720  
 aatttggaaa ctattttgaa agctgccagt ttgatagaga agaataccaa tttacccaag 780  
 aaaattcatt ttcatattgt tggagatggt acggaattgt taagcatgaa agcattagct 840  
 catgaattgg agttaagaa tatttccttc tatggaagac gttctttgga ggaaatgcca 900  
 tcctctata aaaaatcaga tgctatgtta gttcttttaa taggagactc gatagttct 960  
 cgtactatac ctgggaaggt acaatcttat atggcggcag gcaaaccaat tataggtgca 1020  
 atttcaggag atgtcaaaat aattgtagaa gaagcaaatt gtggatatgt tagtcccgaa 1080



```

cgagatgtaa aacaattggc aaaaaatatt tgtaaattta gtatgttatt tattaagaga 1140
caaagagagt taggaaagaa agctcgttgt tactatgaaa atcacttttc aaaagagcag 1200
tttatgctcg aactggagac atgtttagag agggaaagta agaaagaata a 1251

<210> 19
<211> 1128
<212> DNA
<213> Streptococcus pneumoniae

<400> 19
ggaagaagaa tgattaaagt attacatcta tttaaacac tagatagtgg tggagtagaa 60
agttttctat tcaactatta ttctcacatt gatagaaaa aaattcaatt tgattttatt 120
gtgcctggaa aagaacaagg atttttagag gataaaatga aagaattggg tgcaaagggt 180
tatcatgtgc ctctattaag gaaaagcct ctacatcagt ttctctctct tgctagaata 240
ataaagaag gagattatga tatagttcat tgccatggct ataaatctgc aattggtctg 300
atcttatcta aaataattgg ttgtaaaatt agaattattc atagtcata ggcttatgta 360
acagaaaaca gttttcaaaa agtattgcgt aaattagtaa caattttggt aaaaatctta 420
gcaactcatt ggtttgcag tggggaagat tcggctaagt ggttatatgg agagaaagcg 480
tataaagacg gaaaaattga aattattttt aatgcaattg atttgaaaaa gtatcaattt 540
ttgtcagatg ttagagaaaa atgtcgtaga gaattagatg tgtaataaa gttogtatta 600
ggaaatatag ctgcctatc agatcaaaa aaccaaagtt atttatttaa cgttttaaaa 660
gaactcattt taatcaaacc aaatgttatt ttactcctag ttggtaatgg tgaggatgag 720
cagaattaa aacagaagc tttagaacta aatctgacc catatgtgct atttttaggg 780
agaaggactg atatttctga ttattattct cggatggatg tttttttgct tccgtctaaa 840
tatgaggggt tgctctttc tctagtagag gctcaggcat cgggattaca aattttatcg 900
tcagatacag tgacgcaaga agtagatgtg accaaaaaca ttagttactt acctatcaac 960
gaagagtctg tgttgcattg gaaagataaa gtactgtctt taacatctga ggaatgcaat 1020
cgttttgaaa taaataacag tatgacagat ggactctatg atatttgta tcaagctagt 1080
aaattattga atcgttatca agaaatgtgt gtaataaagg agatatag 1128

<210> 20
<211> 1245

```

<212> DNA  
 <213> *Streptococcus pneumoniae*  
  
 <400> 20  
 ttogtgaag ttgatagaat ttcatttata aaaaatacaa gttctctcta tattctgaat 60  
 attgttaaat tactatttcc ttatttaaca ctcccgattt tgacaagggt gotttlogcta 120  
 gacgcgatg gaatggttat ttatgttaaa gogttaatag cttatgttca actgggtgatt 180  
 gattttgggt tcatgatatc agctacaaaa aatattgtaa atgctigtac taactcootca 240  
 aagattggaa ggaatagttgg agatactcta gttgaaaaa tatttttattc tatcattttcg 300  
 attctaattt acaccatatt gatgtggcaa atoccaataa tgagagagaa tattcttttt 360  
 tcagtttttt atttggttagc tacagtgacc aatattttta tctttgactt tttattttcg 420  
 ggaattgaaa agatgcatgc agttgcaatt cottatatta ttctaaaaac tatcattaca 480  
 attttgacat ttattgtagt aaaagatgat tcttctattt tatggattcc tatattggaa 540  
 ggaattggga atttagttgc tgcagtagtt tcttatagat tccttcatta ttatggaatt 600  
 aaattatcat tttcttctc gtctgtttgg gttaaagatt taaaggaatc ctctatttat 660  
 tttttatoca attttgcaac tactattttt ggctgtctta cgacagtcac ttcgggtttt 720  
 tatttacaaa gtcaagagat agccttttgg gggatagcaa tgcaactgct ttcagcagca 780  
 aaatcattgt ataactctat agcgaatagt ttatatccgc atatgatacg tactaaagat 840  
 atacaatcgg ttaagagtat taatcggatt atgtttattc ctattatctt tggagttttg 900  
 atagttttat tcttttcaaa tcaaattctt tctataattg gtggtgaaaa atataccggt 960  
 tcagcagatt tctttaagta cttattaccc gcttttggtg ctagttttta tctatgatgatt 1020  
 tacggatggc ctgtcttagg agctattgat aaagtgaag aaactacaat gacaactata 1080  
 ttagcttoga ttgtccaaac tttgggatta ggaatatata tcttgtctga taatttttagt 1140  
 ttagtaacat tagctatttg ttcaagtatg tctgaggttg tgttatggat tagccggttat 1200  
 ctaatttatt ttaagaacgg ttcattatct gttaggagta agtaa 1245  
  
 <210> 21  
 <211> 5310  
 <212> DNA  
 <213> *Streptococcus pneumoniae*  
  
 <400> 21  
 aagtttatga ataaaggatt atttgaaaaa cgttgtaaat atagtattcg gaaattttca 60

ttaggtgttg cttctgttat gattggagct gcattctttt ggacaagtcc ggttcttgca 120  
 gatagcgtgc agtctgggtc caccggcaac ttaccagctg atttagctac tgctcttgca 180  
 acagcaaaag agaattgatgg gcgtgatttt gaagcgcta aggtgggaga agaccaaggt 240  
 tctccagaag ttacagatgg acctaaagaca gaagaagaac tattagcact tgaaaaaaaa 300  
 aaaccggctg aagaaaaacc aaaagaggat aaacctgcag ctgctaacc tgaaacacct 360  
 aagacggtaa cccctgaatg gcaaacggta gcgaataaag agcaacaggg aacagtcact 420  
 atccgagaag aaaaagggtg ccgctacaac caactatcct caactgctca aaatgataac 480  
 gcaggcaaac cagccctggt tgaaaaagag ggcttgaccg ttgatgccaa tggaaatgca 540  
 actgttgatt taaccttcaa agatgattct gaaaagggca aatcacgctt tgggtctctt 600  
 ttgaaattta aagataccaa gaataatgtt ttgtctggtt atgacaagga tggctgggtc 660  
 tgggagtata aatctcaac aactagcact tggatatagag gtagtctgtt tgctgctcct 720  
 gaaacaggat caacaaaccg tctctctatc actctcaagt cagacgggtc gctaaatgcc 780  
 agcaataatg atgtcaatct ctttgacaca gtgactctac cagctgcggt caatgaccat 840  
 cttaaaaatg agaagaagat tcttctcaag gcgggctctt atgacgatga gcgaacagtt 900  
 gttagcggtta aaacggataa ccaagagggg gtaaaaacag aggatacccc tgctgaaaaa 960  
 gaaacaggtc ctgaagttga tgatagcaag gtgacttatg acacgattca gtctaaaggtc 1020  
 ctcaaacgag tgattgacca agccttccct cgtgtcaagg aatacagctt gaacgggcat 1080  
 actttgccag gacagggtga acagttcaac caagtcttta tcaataacca ccgaatcacc 1140  
 cctgaagtca ctataagaa aatcaatgag acaacagcag agtacttgat gaagcttcgc 1200  
 gatgatgctc acttaataca tgcggaaatg acagtacgct tgcaagttgt agacaatcaa 1260  
 ttgcactttg atgtgactaa gattgtcaac cacaatcaag tcactccagg tcaaaaagatt 1320  
 gatgacgaaa gcaactactt ttctctatt agtttctctg gcaatgcttt agtctctgtt 1380  
 tctagtaatc aaactgggtg taagtttgat ggggcaacca tgtcaaaaca tacgcatgtc 1440  
 agoggagatg atcatatcga tgtaaccaat ccaatgaagg atttggttaa gggttacatg 1500  
 tatggatttg tttctacaga taagcttgct gctgggtgtt ggagtaactc tcaaaaacagc 1560  
 tatggtggtg gttcgaatga ctggactcgt ttgacagctt ataaagaaac agtcggaaat 1620  
 gccaaactatg taggaatcca cagctctgaa tggcaatggg aaaaagctta taaggccatt 1680  
 gtttcccg aatacacgaa ggaacttcca agtgctaagg ttgttatcac tgaagatgcc 1740

aatgcagaca agaactgtga ttggcaagat ggtgccattg cttatcgtag cattatgaac	1800
aatcctcaag gttgggaaaa agttaaggat atcacagctt accgtatcgc gatgaacttt	1860
ggttctcaag cacaaaaacc attccttatg accttggatg gtatcaagaa aatcaatctc	1920
catcacagatg gtcttgggca aggtgttctc cttaaaggat atggtagcga aggccatgac	1980
tctggtcact tgaactatgc tgatattggt aagcgtatcg gtggtgtcga agacttcaag	2040
accctaattg agaaggctaa gaaatatgga gctcatctag gtatccacgt taacgcttca	2100
gaaacttatc ctgagctctaa atacttcaat gaaaaaattc tccgtaagaa tccagatgga	2160
agctatagct atggttggaa ctggctagat caaggatca acattgatgc tgcctatgac	2220
ctagctcatg gtctgttggc acgttgggaa gatttgaaga aaaaacttgg tgacggtctc	2280
gactttatct atgtggacgt ttggggtaat ggtcaatcag gtgataacgg tgcctgggct	2340
acccacgttc ttgctaaaga aattaacaaa caaggctggc gctttgcgat cgagtggggc	2400
catggtgggt agtacgactc taccttccat cactgggcag ctgacttgac ctacggtggc	2460
tacaccaata aaggatcaaa cagtgccatc acccgcttta tccgtaacca ccaaaaagat	2520
gcttgggtag gggactacag aagtattggt ggtgcagcca actatccact gctaggtggc	2580
tacagcatga aagactttga aggtggcag ggaagaagt actacaatgg ctatgtaacc	2640
aacttatttg cccatgacgt catgactaag tacttccaac acttcactgt aagtaaatgg	2700
gaaaatggta caccggtgac tatgaccgat aacggtagca cctataaatg gactccagaa	2760
atgcgagtgg aattggtaga tgctgacaat aataaagtag ttgtaactcg taagtcaaat	2820
gatgtcaata gtccacaata tcgcgaacgt acagtaacgc tcaacggacg tgtcatccaa	2880
gatggttcag cttacttgac tccttgggaa tgggatgcaa atggtaagaa actttctact	2940
gataaggaaa agatgtacta cttcaatacg caggccggtg caacaacttg gacccttcca	3000
agcgattggg caaagagcaa ggtttacctt tacaagctaa ctgaccaagg taagacagaa	3060
gagcaagaac taactgtaaa agatggtaaa attaccctag atcttctagc aaatcaacca	3120
tacgttctct atcgttcogaa acaactaat cctgaaatgt catggagtga aggcattgac	3180
atctatgacc aaggatttaa tagcgttacc ttgaacatt ggaccatttc aggcgatgct	3240
tctaaggcag aaattgtcaa gtctcaagg gcaaacgata tgcttcgtat tcaaggaaac	3300
aaagaaaang ttagtctcac tcagaaatta actggcttga aaccaaatac caagtatgcc	3360

```

gtttatgttg gtgtagataa ccgtagtaat gccaaaggcaa gtatcactgt gaatactgggt 3420
gaaaaagaag tgactactta taccataaag tctctcgcgc tcaactatgt taaggccctac 3480
gcccacaata cagctcgtga caatgctaca gttgacgata caagttactt ccaaaaacatg 3540
tacgccttct ttacaactgg agcggcagtc tcaaatgtta ctctgacatt gagtcgtgaa 3600
gctggtgata aagcaactta ctttgatgaa attcgtacct ttgaaaacaa ttcaagcatg 3660
tacggagaca agcatgatac aggtaaaggc accttcaagc aagactttga aaatgttgct 3720
cagggatatct tccattttgt agtgggtggg gtcgaagggtg ttgaagataa ccgcactcac 3780
ttgtctgaaa aacacaatcc atatacacia cgtgggttga atggtaagaa agtcgatgat 3840
gttatcgaa gaaattggtc actcaagaca aatggactag tgagccgtcg taacttgggt 3900
taccaaacca tcccacaaaa ctccgctttt gaagcaggta agacctaccg tgtaaccttt 3960
gaatacgaa gaggatcaga caatacctat gctttttagt tcggtaaggg agaattccag 4020
tcaggctgctc gtggtaacta agcaagcaac ttggaatgc atgaattgcc aaatacttgg 4080
acagattcta agaaagccaa gaaggcaacc ttctttgtga cagggtgcaga aacaggcgat 4140
acttgggtag gtatctactc aactggaaat gcaagtaata ctctggtgta ttctggtgga 4200
aatgccaaact tccgtgggta taacgacttc atgatggata atcttcaaat cgaagaaatt 4260
acctaacag gtaagatgtt gacagaaaaa gctctgaaga actacttgcc aacgggtgcc 4320
atgactaact acaccaaaaga gtctatggat gctttgaaag aggcgggtctt taacctcagt 4380
caggccgatg atgatatcag tgtggaagaa gcgcgtgcag agattgccaa gattgaagct 4440
ttgaagaatg ctttgggtca gaagaagacg gctttggtag cagatgactt tgcaagtctt 4500
acagctcctg ctcaggctca agaaggtctt gcaaatgcct ttgatggcaa tgtgtctagt 4560
ctatggcata catcttgtaa tgggtgggat gtaggcaagc ctgcaactat ggtcttgaaa 4620
gaaccaactg aaatcacagg acttcgctat gttccgcgtg gatcagggtc aaatggtaac 4680
ttgcgagatg tgaaacttgt tgtgacagat gactctggca aggagcatac ctttactgca 4740
actgattggc caaataacaa caaaccacaa gatattgact ttggtaagac aatcaaggct 4800
aagaaaattg tcttacttgg taccagaca tacggagatg gtggagataa ataccaatct 4860
gcagcggaac ttatctttac tcgtccacag gtagcagaaa cactcttga cttgtcaggc 4920
tatgaagcag ctttgggttaa ggctcagaaa ttaacagaca aagacaatca agaggaagta 4980
gctagcgttc aggcaagcat gaaatatgcg acggataacc atctcttgac ggaagaagatg 5040

```

```

gtggaatact ttgcagatta tctcaaccaa ttaaagatt ctgctacgaa accagatgct 5100
ccaactgtag agaaacctga gtttaaactt agatcttttag ctcccgagca aggtaagacg 5160
ccagattata agcaagaaat agctagacca gaaacacctg acaaatctt gccagcaaca 5220
ggtgagagtc aatctgacac agcctctatc ctagcaagtg ttagtctagc cctatctgct 5280
ctctttgtag taaaaacgaa gaagactag 5310

```

```

<210> 22
<211> 717
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 22
aaggagagg atgaacctat gagaaaattt aaaatctttt tatttatcga agcctgtctt 60
ctgacaggag ctctgatttt gatggtatca gaggcttttt cgcgttttct gctgatacta 120
ttctcttttt tgcttttgat tcgctactac actggtaag agggaaataa tcttctttta 180
gtagcggaac ccattctctt cttttctatc gttatgtcca atccttttgt gattctagct 240
atthttgttg cgggttatcta tagcctcttt cttctttacc cgatgatgaa ccaggaaaaa 300
gagcagacca atthtggttt tgaagaggct gtgacgggta agaaggagaa aaatcggttg 360
tttggaaatc ttcattcatt tccaagctac cagacttgcc aattcgatga tatcaatctc 420
tttcgcttca tgggcaagga cactattcat ctggagaggg tcattctaac caatcatgac 480
aatgtcatta tcttcagaaa gatggttaga acgacaaaa tcactgtacc tgtagatgtg 540
gaagtgcagc tcagcggtta ctgtctctat ggggatttga tttttttcaa ccagcccaag 600
cgagcctctc gcaatgaaca ctatcatcaa gaaacaaaag actatctcaa gagtaacaag 660
agtgtcaaga ttttcttgac cactatgatt ggtgatgtgg aggtgggttag aggatga 717

```

```

<210> 23
<211> 252
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 23
gaggatacaa taatgaagaa aactgtttat aaaaaattgg gtatttcaat tattgcgagt 60
actttattgg ctagccagtt atcgacagta tctgctttga gtgtatttct tagtacaggt 120
gaagaatatg aggttaagtga gacactagaa aaagggtccag agtctaatga ttcttcatta 180

```

tctgagat ttt caccaacgta tgggtcatac taccaaaagc aatcagaagt attatcggta	240
atgatgat ttt ga	252
<210> 24	
<211> 2361	
<212> DNA	
<213> Streptococcus pneumoniae	
<400> 24	
acattcaaa g acaaggaaat aaagatgaat aagaaaatat tagaacatt agagtto gat	60
aagg tcaagg ccttg tttga gcctcatttg ttgaccgagc agggccttgga gcaattgaga	120
caactggctc cgactgccaa agcagataaa atcaaacagg cttttgctga gatgaaggaa	180
atgcaggctc ttttctg tga gcaaccgcat tttactatc tctcaactaa ggaattgca	240
ggagtctgca agaggttgga gatgggagcg gatctcaata tcgaggagtt cctactcttg	300
aaacgcgtgc ttcttgccag ccgagaactt caaaattttt acaccaatct ggaatgtgc	360
agcttggaag aattagccct ttgg ttttgag aaattacatg attttccgca attacaagga	420
aatcttcagg cctttaa tga tgcgggttct attgaaaatt ttgccagtga agaattggcg	480
cgaatccgtc gaaaaataca tgatagcgag agtcaggtag gcgatgtttt acaagacttg	540
ctcaagcaaa aagcgcagct gttgacggaa ggaattgttg ctagcagaaa tggccgtcag	600
gtttttaccag taaaaaacac ctaccgcaat aagattgcag gtgtcgttca tgatatttct	660
gctagtggaa acaccgtcta tatcgaaacc cgtgaggtag tcaaactgag cgaagaaatt	720
gctagtctgc gagcagatga gcgctatgaa atgcttcgca ttctccaaga aatttctgag	780
cgtgtccgcc ctactcggcg tgagattgct aatgacgctt ggattatcgg tcatctggac	840
ttgattcgtg ccaaggttcg attttacaa gaaagacaag cagtcgtgcc tcagctgtca	900
gaaaaatcaag agattcaact gctccatgtc tggcatcctt tggtaaaaa tggcgtcgca	960
aatgatgtct attttgg tca agatttaaca gctattgtca ttacaggctc caatacagg	1020
gggaagacca tcatgctcaa aactctgggc ttgacacagg tcatggccca gtcaggattg	1080
ccgatttttag cagacaaggg aagtcgtgtt ggtatttttg aagaaatcct tgctgatatt	1140
ggagatgagc agtctattga gcagagcttg tctaccttct ctagtcatat gaccaatctc	1200
gtggatattc ttggcaagg t caaccaatc tctactctac ttttggatga gttgggggct	1260
ggtactgatc cccaagaggg agcagccctt gccatggcta ttctggagga ccttcgcctg	1320

```

cgTcaaatca agaccatggc gacgaccac tatccagaac tcaaggccta cggTattgag 1380
acagcctttg tgcaaaatgc cagtatggag tttgatactg caactcttcg cccgacctat 1440
cgctttatgc aggggtgttc tggccgaagt aatgcctttg aaattgcaa acgtctaggc 1500
ctatctgaag ttatcgtagg agatgccagt cagcagatcg atcaggacaa tgacgtcaat 1560
cgtatcattg agcaattaga agagcagacg ctggaaagcc gcaaacgttt ggaacaatgc 1620
cgtgaggtgg agcaagaaaa tctcaagatg aaccgtgcgc taaaaaact ctacaacgag 1680
cttaalcgtg aaaaggaaac cgagcttaac aaggcgcgtg aacaggctgc tgagattgtg 1740
gatatggccc taagtgaag tgaccagatt ctcaaaatc tccacagtaa atcccaactc 1800
aagccccacg aaatcattga agccaagccc aagttgaaaa aattggctcc tgaaaaagtg 1860
gacttgctca aaaataaggt ccttcaaaag gccaagaaaa aacgagctcc aaaggtggga 1920
gatgatatcg tggttctcag ttatggtcag cgtggtacct tgaccagtca actcaaggac 1980
ggtcgctggg aagcccaagt tggcttgatt aagatgacct tggaagagaa agagtttgat 2040
cttgttcaag cccagcaaga aaaaccagtc aagaagaaac aggtcaatgt tgtgaaacga 2100
actttcgggc gaggacctca agctagactg gatcttcgag gcaagcgcta tgaagaagcc 2160
atgaatgagc tagatacctt catcgacca gcttgcctta acaatatggc tcaagttgat 2220
atcatccatg gtatcggaac aggagtcacg cgtgaaggag ttaccaaata cttgcaaaga 2280
aacaacatg tcaagagttt cggctatgcc ccacaaaatg ctggaggcag tgggtgcgact 2340
attgtcactt ttaaggata g 2361

```

```

<210> 25
<211> 294
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 25
cagctgaggc acgactgctt gtctttcttg gataaatcga accttggcac gaatcaagtc 60
cagatgaccg ataattcaag cgtcattagc aatctcagcc gcatgagggc ggacacgctc 120
agaaatttct tggagaatgc gaagcatttc atagcgctca tctgctcgca gactagcaat 180
ttcttcgctc agtttgacta cctcacgggg ttcgatatag acggtgtttc cactagcaga 240
aatatcatga acgacacctg caatcttatt gcggtagggt tttttgactg gtaa 294

```

```

<210> 26

```



&lt;211&gt; 915

&lt;212&gt; DNA

<213> *Streptococcus pneumoniae*

&lt;400&gt; 26

ttattggagg ttaggatgaa aaaactcccc ttagtatctt caggttggtt gctaggtttg	60
gcaggagctg gaaatcttat tttagatacg ttgccgggtc tatcccatct ttttagtctg	120
attgggttgg ttttatggat ttactttcta attctgcac tctttaattg gaaagaaacc	180
aagcaagaat tgaccaagcc ccctcttttg tcagggaatgg caacctttcc tatggctggg	240
atgattttat cgacctatgt ctttcgcgta ttctcttacc ttcttttggt agcacaaggg	300
atttggtggt tttcatttct cttggatttg accttgattg ctggttttac catcaagttt	360
gcttgccag ggcgggggtg tcatgccact ccaagctgga cggttctcta tgtggggata	420
gcagtggctg ccttgacctc tcctctgcta ggtattatcg aaattgccta tgcgaccttg	480
agttttgggt ttctcttgac cttctatctc tatccctta tttatagcga tttaaagaaa	540
catccactcc cactagcctt gcttgacaaa gaaggaatct actgtgctcc tttctctcta	600
ctcttggtct ctctagttcg agtaggagga accagcctgc cgacttggtt cttgattgtc	660
atgatttttg cttctcaatc cttcttttct tttgttttaa ctgctctgcc caacatttta	720
aaacaagggt ttcaaccagc cttctcagcc ctcaccttcc caaccattat cacagcgacc	780
tcgctcaaga tggctcaggg aattttgaaa cttccatttc tggattacct ggtattggct	840
gaaaccatta tatgcctaac tattttatct tttgtactag gtgcttatct gatttggtta	900
cgaaaaaagg tctag	915

&lt;210&gt; 27

&lt;211&gt; 849

&lt;212&gt; DNA

<213> *Streptococcus pneumoniae*

&lt;400&gt; 27

tctatgtatc ttattgaaat tttaaaatct atcttcttcg gaattgttga aggaattacg	60
gaatgggttc cgatttccag tacagggtcac ttgattttag cagaggaatt catccaatac	120
caaaatcaaa atgaagcctt tatgtccatg tttaatgtcg tgattcagct tgggtgtatt	180
ttagcagtta tgggtgattta ttttaacaag ctcaatcctt ttaaacggac caaggacaaa	240
caggaagttc gtaagacttg gagactatgg ttgaaggctc tgattgtcac tttaccttta	300
cttggtgtct ttaaatgtga tgattgggtt gataccact tccataacat ggtttcagtt	360

gctctcatgt tgattatcta cgggggtgcc ttcattctatt tggaaaagcg caataaagcg 420  
 cgtgctatcg agccaagtgt aacagagttg gacaagcttc cttatcagac cgctttctat 480  
 atcggactct tccaagttct tgctctttta ccagggaacta gccgttcagg tgcaacgatt 540  
 gtcggtgggt tggttaaatgg aaccagtcgt tcagttgtga cagaatttac cttctatctt 600  
 gggattctctg ttatgttttg agctagtgc ttaagattt tcaatttgt gaaagccgga 660  
 gaactcttga gctttgggca attgttttg ctcttggtcg cgatgggagt agcttttgcg 720  
 gtcagcatgg tggctattcg cttcttgacc agctatgtga aaaaacacga cttcacctt 780  
 tttggttaat accgtatcgt gcttggtagt gtttgctac ttacagttt tgcctgttta 840  
 tttgtataa 849

<210> 28

<211> 939

<212> DNA

<213> *Streptococcus pneumoniae*

<400> 28

aatgatgagt ttgaagataa agggatgctg ataaaatgg taaaacaaa aaagcaaaaa 60  
 cgaaataatc tctatttagg agtggatttt ttcattggaa tggcggtaat ggcgtatccg 120  
 ctgggtgtctc gcttgatta tcgagtggaa tcaaatcaac aaattgctga ctttgataag 180  
 gaaaaagcaa cgttggatga ggctgacatt gatgaacgaa tgaaattggc acaagccttc 240  
 aatgactctt tgaataatgt agtgagtggc gatccttggt cggaagaaat gaagaaaaaa 300  
 gggcgagcag agtatgcacg tatgttagaa atccatgagc ggatggggca tgtggaaatc 360  
 ccggtatttg acgtggattt gccggtttat gctggtagct ctgaagaggt attgcagcaa 420  
 ggggctgggc atctagaggg aacttctctg ccgacggag gcaattcgac ccatcggtg 480  
 attacggcac atacaggttt gccaacagct aagatgttta cggatttgac caaacttaaa 540  
 gttggggata agttttatgt gcacaatatc aaggaaatga tggcctatca agtggatcaa 600  
 gttaaaggta ttgagcgcag gaactttgat gattttatga ttgtaccagc tcatgattat 660  
 gtgaccttgc tgacttgtac gccatacatg atcaataacc atcgtctatt ggttcggggg 720  
 catcggatag cgtacgtagc agaggttgag gaagaattta ttgcagcaaa caaactcagt 780  
 catctctatc gctacctgtt ttatgtggca gttggtttga ttgtattct tttatggatt 840  
 attcgacgct tgcgcaagaa gaaaaacaa ccggaaaagg ctttgaaggc gctgaaagca 900

gcaagggaagg aagtgaaggt ggaggatgga caacagtag 939

<210> 29

<211> 903

<212> DNA

<213> *Streptococcus pneumoniae*

<400> 29

agggtggagga tggacaacag tagacgttca cgaaaaaag gcacaaaaa gaagaacat 60

ccgctgatcc ttcttctgat ttctcttagta ggattcgcgc ttgcgatata tccattgggtg 120

tctcgttatt attatcgtat tgagtcaaac gaggttatta aagagtttga tgagacgggt 180

tcccagatgg ataaggcaga acttgaggag cgttggcgct tggctcaagc cttcaatgcg 240

accttgaaac catctgaaat tcttgatcct tttacagagc aagagaaaaa gaaaggcgctc 300

tcagaatatg ccaatatgct aaagggtccat gagcggattg gctatgtgga aattcctgcg 360

attgatcagg aaattccgat gtatgtcggg acgagtgagg acattcttca gaaaggggca 420

gggctgttag aaggggcttc gctgcctgtt ggagtgaaa ataccatac agtgatcact 480

gctcacagag gattgccaac ggcagaattg ttcagtcaat tggataagat gaaaaagggg 540

gatatctttt atcttcacgt tttagatcag gtgttggcct accaagtgga tcagatagtg 600

acgggtggagc cgaatgactt tgagcctgtc ttgattcaac atggggaaga ttatgcgacc 660

ttgttgactt gtacacogta tatgattaac agtcatcgctc tgttggtacg tgggaagcgg 720

attcogtata cggcaccaat tgcagagcgg aatcgagcgg tgagagagcg tgggcaattc 780

tggttgtggt tattactagg agcagatggcg gtcaccccttc tcttgctgta tcgcgtgtat 840

cgtaatcgac ggattgtcaa aggactagaa aagcaattgg aggggcgtca tgtcaaggac 900

taa 903

<210> 30

<211> 1347

<212> DNA

<213> *Streptococcus pneumoniae*

<400> 30

aaaataaaaa aaggagttcc ggtgatgaac aatatttttag cgttttttaga aacaaaagtc 60

gctccgtttg gtgaaaaagt tggcaaccaa cggcattttga aagctattcg tgaaggattt 120

atgatggcna tgccctttgat tttagtcggc totttatttc ttattcctaat cagttggcct 180

```

caagaggcctt ttacaaattg gctgaatagt gttggattgc taagtatcctt gacaactatg 240
aatcagtgcaa cagtagcgat tatctccttg gtcgcttggt tcggtattgc ctacagggtg 300
tcggaaggat atggtacaga tggccgctcg gcagggatca tagccttate cagttttgta 360
ttgatggcac ctctgttttc gagtatgggt tatgataaaa atggggagca ggtcaagcag 420
ttatttggcg ggcgaatacc atttcttagc ctgaatgcct ctctttggtt tatggcgatt 480
actattggat tgggtacagc agagatttat cgtatgttta tccagcgagg aattacgata 540
aaaatgccaa gtggtgtccc agatgtagta agtaaatcat ttacagctct ttatctgggt 600
ttactactct ttgttttggt ggctttgggt ttaaaaggct ttgaagcggc aggagtgtca 660
ggaggtctca acgactcctt aggtgcaatt gttggaacac cgcttaagtt aattgcagga 720
acgcttcacg gtatgattct atgtgttatt gtaaatcat tcttttgggt ctgtggaggt 780
aatgggggac aagttttaaa tgcctttgta gaccagttt ggttacaatt tactacagaa 840
aaccaagaag ctgtggctgc aggacaaaca ctccaacaca ttattacatt accgtttaaa 900
gatttatttg tatttattgg tggcggtgga gcgactattg gtcttgcat ttgtctcttc 960
ctatttagta agagtctgac gaataaaaca ttaggtaagc tagctattat accgtctatt 1020
ttaatatca atacagctat tctatttacg ttccaacag ttttaaatcc gattatgctg 1080
attccgttta ttgctactcc tacaatcaat gccttgatta cctatgtatc aatggctgta 1140
ggatttagtac cctatacaac aggtgtaatc ctccgtgga caatgccacc gattatagga 1200
ggcttccttg caacaggggc tagttggcga ggagctctat tacaagttgt tttgattttg 1260
gtttctgtag caatttatta tccattcttc aaaattgcag ataaacgcaa tcttgaaaaa 1320
gaaaagcta ctgttgagg gaaataa 1347

```

```

<210> 31
<211> 1701
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 31
attttttata ggaggagttt tatggataag ctagtgcgtg ccattgaaa gcaacaaggg 60
aaatttgaaa aaatttttac taataactat atgatggcta ttaaagatgg attcattgct 120
actatgcctt taattatggt ttcaagcttt ttgatgatta ttattatgat tcttaaaaaa 180
ttcggagtag agttaccgag tccagctatt gtctggatga gaaaagtga tatgttaacc 240

```

```

atgggagttt tgggtattat tgtttcaggg actgttgga agtcattagt tggaaatggt 300
aacagaaaaa tgccctcacgg aaaggaata aatgatatct ctgcaatggt ggcagccata 360
tgtagttatc tggtagtaac tgtaacgctt gtagttgatg agaagacggg atctacaagt 420
ttgtcgacaa actatttagg atctcaagga ttgataactt cgtttgtcag tgcctttatt 480
actgtaaatg tttaccgatt ctgtattaag cgagacatta ctattcattt acctaaaggaa 540
gttcctgggg ctatatcaca agctttttaga gataatttcc ctttttcttt tgttttactt 600
attagtggtt tgtagatat tgtatctcgg tttagtttag atgttccttt tgcccaagta 660
tttcaacaac tattgactcc tatttttaag ggggcagaat catatcctgc tatgatgttg 720
atttggttta tgtgtgcttt gctttggttt gttggaattc atggaccatc tattgtctta 780
cctgctgta cagctttgca actgagcaat atggaagaga atgctcaact tcttgcaa 840
gggcagtctc cttatcatcc ttaaacacct aatttcggga attatatcgc tgctattgga 900
ggaaocgggg ctacccttgt tgtaccattt attttgattt tctttatgcg gtctaaca 960
ttaaaatcgg taggtaaagc tacaattact cctgttttat ttgcggtaaa tgaacctctt 1020
ctatttggtt tgctggttat tttgaatccc tatctttttt tccctttttt gatgactcca 1080
ccagtgaatg tatttctagg aaaggctctt attgatttct ttggaatgaa tggattttat 1140
atccagttac cttggacctt tcctggctcc ttgggattgt taattggaac gaattttcaa 1200
cttatctcct ttgtattttt atctttgatt ttagttgtcg acatattgat ttatttgcca 1260
ttctgtagag cgtatgatag acagttaact gtgaaagaag atattgcaag ctcaaatgat 1320
attatttttag aggaggatag aagtgaata attcctggtg agatagatga aataaaaagt 1380
aaggagtgtg aagtactggt tctttgtgca gggctcggaa caagtgcga attagccaat 1440
gcaattaaacg aggggggctaa ctaacacag gttagagtga ttgcgaattc aggagcgtac 1500
ggagctcatt atgatatatt ggggttttat gatttaatta ttctggcccc acaagttcgg 1560
agttattata gagagatgaa ggtggatgca gaaagattag gtattcagat agttgtctacc 1620
agaggaatgg aatatattca ttaacaaag agtccaagta aagccttaca atttgatttg 1680
gagcattacc aagctgtgta g 1701

```

```

<210> 32
<211> 1704
<212> DNA
<213> Streptococcus pneumoniae

```

<400> 32  
 gaagcgaatg aagagagtaa gatgaagaa gctataattg agtgaagga tttctcttct 60  
 cggatatgaaa cacaacaaga accgaccttg caagggatag acttgaccat ttacaagggga 120  
 gagaaagtct taattgttgg accatctggg tcaggtaaat ctaccttggg tcagtggttg 180  
 aatgggatta ttccaatat ttacaagggt cagacatatg gagaattttt gataaagggt 240  
 caagtagcct ttgatatgag catctatgat aagtctcacc tggttagcac agttttgcag 300  
 gatacagatg ggcagtttat tggcttgtct gtggcagaag atttggcggt tgctctggaa 360  
 aatgatgtga cagccctaga tgagatgaaa ggtcgtgttt ataaatgggc tgaaaagctg 420  
 gaccttcttc ctttactgga tcagcgtcct caggatttgc cagggtggaca aaagcagcga 480  
 gtcagtctgg ctggtgtcct gattgatgaa agtccgattc tctgttttga tgagccactc 540  
 gccaatctag atcccaagtc aggtcaggat attatcgaat tgattgacca gattcataag 600  
 gaagagggga cgaagactct tattatcgag cacggtttgg aggacgttct gcacgccct 660  
 gtggatcgga ttgtcttgat aaacgatggt cgtatccttt ttaatgggag ccctgaccag 720  
 ttgcttgcga ctgatttatt gactcaaaat ggaattcgag aacccttcta tctaagcact 780  
 ctccgtcaat tagtgttga cttagtcaag gaagaacaat tagcgaatct ggataacttg 840  
 tctatctcaa aaggtcaggt tcagttgcag aatgaactgg caaaagaaac ccagcattg 900  
 cagtactct ttgactaga ggaagtatct ttttcttatg atgatatgacc gattttaaaa 960  
 tcctacatt tagatattaa aaagggtgaa aagattgcta ttgtcggaaa aaatggagca 1020  
 gggaaatcaa ctctagccaa ggcataagt agctttattc agacggaagg acgctatctt 1080  
 tgggaaaaac aggataataa aggcgattct gttgcagagc gggcggaaac agtaggatat 1140  
 gtgtacaaa atcctaataa aatgatttca accaatatga tttttgatga ggtggctcta 1200  
 gggctccggt tgcgaggtgt ggatgagaag gaaattgaaa cgagagtata tgaaaccttg 1260  
 aaaatctgtg gactttatga attccgtaat tggcctattt ctgccctgtc atttggtcag 1320  
 aaaaaactgt tcaccattgc ttcaattttg gtcttaggag ctgaaattat tctctagat 1380  
 gaaccgactg caggtcaaga tcagaagaac tatactgaga ttatggaatt tctcgaagag 1440  
 ttacatcaaa aagggcatac cattgtcatg attacccatg atatgcaatt gatgctggat 1500  
 tattcagacc gggctctgt catggtggat ggagaattga ttgccgatac tgttccagcc 1560  
 agtctgttga gcgactctga gctgttagta aaagccaatc taaaagaaac ctccactctt 1620

aacttggcta agaaactaga tgtggatcca ctggatttaa cggcatttta caaagaaagg 1680  
 agagaggggat gcaagctaaa ttaa 1704

<210> 33  
 <211> 1668  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 33  
 gagatattcc gcgattattt ttcggaggaa aaagaaatgt ccattaattg gcaggaaatt 60  
 ttatttcact ttttaggttg tctggggcta ttcttatata gcatcaagac catgggagac 120  
 ggtttacaac aagctgctgg agatcgctt cgtttttaca ttgacaaata tactagtaaat 180  
 cctttctttg gagttctggt tggatttggg atgactgctc taattcagtc tagttctggt 240  
 gtaacagtta tcacagtcgg cctggtcagt gccggtctct taaccttacy tcaggctatc 300  
 gggattgtca tgggtgctaa tattgggaca actgtcacat cctttctcat cggtttttaa 360  
 ttaggttaact atgcccctacc tatgctcttt atcgggtgcg tctgtctttt ttttacgaaa 420  
 aatcggacag tcaataatat cggacgcato ctctttggtg tcggtggtat cttttttgcc 480  
 ctcaatctca tgagcggcgc aatggctcca ctcaaggatt tacaggctct taaggactat 540  
 atgattgagc taagtaagaa tcctgttttg ggtgtctttg tcggtactgg cttgaccttg 600  
 ctaattcaag cttcttggc taccattggg attttacaaa acctctacgc cggcaatcta 660  
 attgatctac agggagcttt gccagttcta tttggtgaca atatcgggac aaccattaca 720  
 gccatcattg cctcttttag ggctaataatt gcagctaaac gggtagcagg agctcatggt 780  
 gccttcaacg ttatcggaa agttgtctgc gttatttttc tagttccttt tactgtcctg 840  
 attcattggt ttgaagctac gctaaatcta gcaccggaaa tgaccatgc cttgtctcac 900  
 ggaaccttta atattacaa caccattgtc caatttccat ttatcggagc tctggtctac 960  
 tttgtaacca agattattcc tggagaggac gaggttgtca aatacgaacc cttatatctt 1020  
 gatgaacatt tcatcaaca gggcccatct atcgctctag gaaatgctaa gaaagagctc 1080  
 ttgcacttag gaaactacgc tgcataagcc tttagccttt cctataagta catcattgac 1140  
 ttggatgaaa aagttgctga aaaagggcat aaaaccgaag aagcaattaa caccatcgat 1200  
 gagcaattaa cacgttatct cattgccctt tcaagcgaag ctctcagcca aaaagaaagt 1260  
 gaagtgttta ccaatatcct tgattcctcc cgtgatttgg aacggattgg agaccacacg 1320

gaggctctac tcaatctgac tgactatctt caacggaaaa atgttgaatt ttctgatgcc 1380  
gccttgaaag aattagagga agtttacgc caaactagtg actttatcaa agatgctctg 1440  
gatagtgtgg aaaacaatga tattgaaaa gcacgcagtc ttgtagaacg tcatgaagca 1500  
atcaataaga tagaacgtgt tctcagaaaa acccacatca aacgcctcaa caaaggcgaa 1560  
tgttcaacac aagctgggtt caactttatc gacatcatct cactacacac togtgtatca 1620  
gaccacgcta tgaaccttgc tgaaaagggt ttgcagaac aaatctaa 1668

<210> 34  
<211> 4989  
<212> DNA  
<213> Streptococcus pneumoniae

<400> 34  
gaggagaaaa tgaagaatcc attttttgaa agacgttgct gttacagtat tcgtaagtta 60  
tcagtaggag cctgctcgtc gatgattggt gctgttttat ttgctgggtc agccttggct 120  
gaagaaactg cagttctctga aaatagcgga gctaatacag agcttggttc agggagagagt 180  
gagcattcga ccaatgaagc tgataagcag aatgaagggg aacatgctag agaaaacaag 240  
ctagaaaagg cagaaggagt agcgtatgca tctgaaactg cttcgccagc aagcaatgaa 300  
gctgcaacta ctgaaactgc agaagcagct agcgcagcta aaccagagga aaaagcaagt 360  
gagggtggtg cagaaacacc atctgcagaa gcaaaacctt agtctgacaa ggaacacaaa 420  
gcaaagcccg aagcaactaa ccaaggggat gagtctaaac cagcagcaga agctaataag 480  
actgaaaag aagtcacgcc agatgtccct aaaaatacag aaaaacatt aaaaccasag 540  
gaaatcaaat ttaattcttg ggaagaattg ttaaaatggg aaccaggtgc tcgtgaagat 600  
gatgctatta acccgcgatc tgttgtcttc gcttcacgtc ggacaggtca tttagtcaat 660  
gaaaagccta gcaaggaagc aaaagttcaa gccttatcaa acaccaattc taaagcaaaa 720  
gaccatgctt ctgttggttg agaagagttc aaggcctatg cttttgacta ttggcaatat 780  
ctagattcaa tgggtctctg ggaaggtctc gtaccaactc ctgacgttat tgatgcaggt 840  
cacgtaacg ggggtctctg atacgggtaca ctcttcttca actggtctaa tagtattgca 900  
gatcaagaaa gatttgctga agctttgaag caagacgcag atggtagctt cccaattgcc 960  
cgtaaatggt tagacatggc caagtattat ggctatgatg gctatttcat caaccaagaa 1020  
acaactggag atttggttaa acctcttgga gaaaagatgc gccagtttat gctctatagc 1080



aaggaatatg	ctgctaaggt	aaaccatcca	atcaagtatt	cttggtagca	tgccatgacc	1140
tataactatg	gacgttatca	tcaagatggg	ttgggagaat	acaactacca	attcatgcaa	1200
ccagaaggag	ataaggttcc	ggcagataac	ttctttgcta	actttaactg	ggataaggct	1260
aaaaatgatt	acactattgc	aactgccaac	tggattggtc	gtaatcccta	tgatgtat	1320
gcaggttttg	aattgcaaca	gggtggttcc	tacaagacaa	aggttaagt	gaatgacatt	1380
ttagacgaaa	atgggaat	gcgcctttct	cttggtttat	ttgcccaga	taccattaca	1440
agtttaggaa	aaactgggtg	agattatcat	aaaaatgaag	atatcttctt	tacaggttat	1500
caaggagacc	ctactggcca	aaaaccagg	gacaaagatt	ggtaggtat	tgctaaccta	1560
gttgcggacc	gtacgccagc	ggtaggta	acttttacta	cttcttttaa	tacaggtcat	1620
ggtaaaaaat	ggttcgtaga	tgtaagggt	tctaaggatt	ctgagtggaa	ttatcgttca	1680
gtatcaggtg	ttcttccaac	atggcgctgg	tggcagactt	caacagggga	aaactctcgt	1740
gcagaatatg	attttacaga	tgctataat	ggcggaat	cccttaaatt	ctctgggtgat	1800
gtagccggta	agacagatca	ggatgtgaga	ctttattcta	ctaagttaga	agtaactgag	1860
aagaccaaac	ttcgtgttgc	ccacaaggga	ggaaaagggt	ctaaagttta	tatggcatc	1920
tctacaactc	cagactacaa	attcgatgat	gcagatgc	ggaaagagct	aacctttct	1980
gacaactgga	caaatgaaga	atttgatctt	agctcactag	cgggtaaaac	catctatgca	2040
gtcaactat	ttttcgagca	tgaagggtct	gtaaaagatt	atcagtttaa	cctaggacaa	2100
ttactatct	cggacaatca	ccaagagcca	caatcgccga	caagcttttc	tgtagtgaaa	2160
caatctctta	aaaatgccca	agaagcggaa	gcagttgtgc	aatttaaagg	caacaaggat	2220
gcagatttct	atgaagttta	tgaaaaagat	ggagacagct	ggaaattact	aactggctca	2280
tcttctacaa	ctatttatct	acccaaagtt	agccgctcag	caagtgtc	gggtacaa	2340
caagaactga	aggttgtagc	agtcggtaaa	aatggagttc	gttcagaagc	tgcaaccaca	2400
acctttgatt	gggtgatgac	tgtaaaagat	accagcctac	caaaaccact	agctgaaat	2460
atcgttccag	gtgcaacagt	tattgatagt	actttcccta	agactgaagg	tggagaagg	2520
attgaaggtg	tggtgaacgg	taccattact	agcttgtcag	ataaatggtc	ttcagctcag	2580
ttgagtggtg	gtgtggatat	tcgtttgacc	aagccacgta	ccgttgtag	atgggtcatg	2640
gatcatgcag	gagctgggtg	tgagctctgt	aacgatggct	tgatgaacac	taaagacttt	2700

gaccctttatt ataaagatgc agatggtgag tggaagctag ctaaggaagt ccgtgggtaac 2760  
 aaagcacacg tgacagatat cactcttgat aaaccaatca ctgctcaaga ctggcgcttg 2820  
 aatggtgtca cttctgacaa tggaactcca tggaaggcta ttcgtatcta taactggaaa 2880  
 atgtatgaaa agcttgatac tgagagtgtc aatattccga tggccaaggc tgcagcccg 2940  
 tctctaggca ataacaaggt acaagttggc ttctcagatg taccggctgg agcaactatt 3000  
 accgtttatg ataactcaaa ttctcaaat ccgctcgcaa ccttgaagag cgaagttgga 3060  
 ggagacctag caagtgcacc attggatttg acaaatcaat ctggtcttct ttattatcgt 3120  
 acccagttgc caggcaaggga aattagtaat gtcctagcag ttccggttcc aaaagatgac 3180  
 agaagaatca agtcagtcag cctagaaaca ggacctaaga aaacaagcta cgccgaaggg 3240  
 gaggatttgg acccttagagg tgggtgtctt cgagtccagt atgaaggagg aactgaggac 3300  
 gaactcattc gcctaactca cgcaggtgta tcagtatcag gttttgatac gcatcataag 3360  
 ggagaacaga atcttactct ccaatatatt ggacaaccgg taaatgctaa tttgtcagt 3420  
 actgtcactg gccaaagcga agcaagtcgg aaaactattt tgggaattga agtaagttag 3480  
 gaaccgaaaa aagattacct agttggtgat agcttagact tgtctgaagg acgctttgca 3540  
 gtggcttata gcaatgacac catggaagaa cattccttta ctgatgagg agttgaaatt 3600  
 tctggttacg atgctcaaaa gactggctgt caaaccttga cgcttcatta ccaaggccat 3660  
 gaagttagct ttgatgtttt ggtatctoca aaagcagcat tgaacgatga gtacctcaaa 3720  
 caaaaattag cagaagttag agctgctaag aacaagggtg tctataactt tgcttcata 3780  
 gaagtaaaag aagccttctt gaaagcaatt gaagcggcgg aacaagtggt gaaagaccat 3840  
 gaaactagca cccaagatca agtcaatgac cgacttaata aattgacaga agctcataaa 3900  
 gctctgaatg gtcaagagaa atttcggaa gaaagacag agcttgatcg cttacagggt 3960  
 gagggtcaag aactcttggc tgccaaacca aaccatcctt caggttctgc ctagctccg 4020  
 cttctgaga aaaacaaggc cttggttgaa aaagtagatt tgagtccaga agagcttaca 4080  
 acagogaac agagtctaaa agatctggtt gctttattga aagaagacaa gccagcagtc 4140  
 tttctgata gtaaaacagg tgttgagta cacttctcaa ataaagagaa gactgtcatc 4200  
 aagggtttga aagtagagcg tgtcaagca agtgctgaag agaagaaata ctttctgga 4260  
 gaagatgctc atgtctttga aatagaaggt ttggatgaaa aaggtcaaga tgttgatctc 4320  
 tcttatgctt ctattgtgaa aatcccaatt gaaaaagata agaaagttaa gaaagtattt 4380

ttettacctg aaggcaaga ggcagtagaa ttggcttttg aacaaacgga tagtcatgtt 4440  
 atctttacag caoctcactt tactcattat gcctttgttt atgaatctgc tgaaaaacca 4500  
 caacctgcta aaccagcacc acaaaacaca gtccctccaa aacctactta tcaaccgact 4560  
 tctgatcaac aaaaggctcc taaattggaa gttcaagagg aaaagggtgc ctttcatcgt 4620  
 caagagcatg aaaatactga gatgctatgt ggggaacaac gagtcatcat acagggagca 4680  
 gatggactgt taagacatgt ctttgaagtt gatgaaaacg gtcagcgtcg tcttcgttca 4740  
 acagaagtca tccaagaagc gattccagaa attgttgaaa ttggaacaaa agtaaaaaca 4800  
 gtaccagcag tagtagctac acaggaaaaa ccagctcaaa atacagcagt taaatcagaa 4860  
 gaagcaagca aacaattgcc aaatacagga acagctgatg ctaatgaagc cctaatagca 4920  
 ggcttagcca gccttgggtc tgctagttaa gccttgacct tgagacggaa aagagaagat 4980  
 aaagattaa 4989

<210> 35  
 <211> 1029  
 <212> DNA  
 <213> *Streptococcus pneumoniae*

<400> 35  
 gcaagcttcc ttctctgat tttcaacaaa aaatctctca ttgtttacat tgttctctca 60  
 agcttattgg tcaactattat caatataggt ggttcttact atctccaagg aatcttggat 120  
 gaatacatc caaatcagat gaaatcaact ttaggaatca tctcagttgg tctgggtatc 180  
 acctatatcc tccaacaagt catgagcttc tccagagatt atctcctaac cgttctgagt 240  
 cagagattaa gtattgatgt gatttttatcc tatattcgcc atatttttga acttcccatg 300  
 tctttctttg cgacacgtcg tacaggagaa atcatttcac gattcacaga tgctaactct 360  
 attatagatg cottggcttc taccattctt tctcttttct tggatgttct tattctgatt 420  
 cttgtaggag gcgtcttact ggcacaaaaa cctaactctct tccttcttct tcttatttcc 480  
 attcctatat acatgttcat catcttttct tttatgaaac ctttcgaaaa aatgaaccat 540  
 gatgtoatgc aaagtaattc tatggtttagc tctgccatta tcgaagatat caacgggatt 600  
 gaaactataa agtgcgtcac gagtgaagaa aatgcgtatc aaaatataga cagcgaattt 660  
 gtagattatt tggaaaaatc ctttaagctc agtaaatatt ctattttaca aacgagttta 720  
 aagcagggaa caaaattagt tctgaatatc cttatcctat ggtttggcgc tcaattagtc 780

```

atgtcaagta aaatttctat cggtcagctg attaccctta acacactttt ttcttacttt      840
acaactccta tggaaaaatat tatcaacctc caaaccaaac tccaatctgc gaaggctcgt      900
aataaccgtt tgaacgaagt ctatctagtc gaatctgaat ttcaagttca agaaaacctt      960
gttcattcac attttttgat gggcgatatt gaatttgatg acctttctta taagtatggt     1020
tttggatga                                     1029

```

```

<210> 36
<211> 288
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 36
ggtcttgggg taaaaaaca aaaggcttgc ttttcagcca tagaggaggt catcatgtat      60
aaacacttat ttttctaga ttccaaaact ttgatcgggt tgacacccta tattctagtc     120
ttggcttctg acaccattgc ttttaatggt ttgtgtctaa cctttgtatc tgcgggtggt     180
tttaatttcc taaattccat gctagcttta atggctatat tcataggggc tggctatgtg     240
gtcggatttt ggttactaat actcaatgaa aatcaaagag caaactag                     288

```

```

<210> 37
<211> 648
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 37
cgtgtaggaa gctcgtttgt tgaggaggat aattttatgg agttttttga taaatttcat      60
gccttgtgtt ttgatttttt agtactaata attgtcatta cagttcctta tacgattaac     120
catggggggt tttttcaaaa tgaatctgca ttgattcttg taagtcttct tgtaacctcg     180
ctgagtgttg cttatgctag aaagtttgaa atgatttctt ttgggatggt aagcaagaaa     240
caacttttgc ttttcattgc aatctttctt ctaagtgtac ttgagacgct ggttttatatt     300
catttctctg ctgtttcttc tggctcaggg gtccaacact tggcgggaagt cagcagagga     360
atttcctctg ctttgatttt gactacctca gtttttggcc ccatccagga ggaactcatt     420
ttcagaggac ttcttcaagg tgcgggtttt gacaattctt ggtaggggct tgtgtctaact     480
tcctctctct ttttttctat gcattggacct tctaattgtc cttcgtttat tttttatcta     540
cttggggggt tgttctctgg ctttgcttat aaaaagagtc aaaacctatg ggtttctact     600

```

ctagtcata tgctttacaa cagttggcca ctcttatatt atttataa 648

<210> 38  
 <211> 1848  
 <212> DNA  
 <213> *Streptococcus pneumoniae*

<400> 38  
 gagaatacca tgagttataa agatacggta caaaaaatcc tcgatgtaat tggaggtgaa 60  
 aaaaatgtca atagagttac ccattgtgta acacgtttta gattagaatt aaaagatgaa 120  
 aatttagtca atgatgatga tgtgaagaag ataccaggtg taatagggtat tatgaaaaag 180  
 aatggacaat atcaaattat acttggtaat gatgtagcta attattataa agaattcgtt 240  
 aaacttggca attttgaatc cgattcagtt gttcaagggc acaaagggaa tattttgaa 300  
 agaatcattg agtatatcgc tggttccatg actccaatca ttccagcaat gttaggggga 360  
 ggtatgttga aagctcttgg aatcatttta ccaatgcttg gtatattgca atcagattct 420  
 cagactattg cttttttgac attttttggg gatgtccat attatttctt accgctgtta 480  
 ttagcttatt ctgcatcaca aaaattaaaa gtaacatcta cattagctat gtctgtagca 540  
 ggtgtacttc tccatccaaa ttttgttcaa atggtgcaat cagggaaatcc tcttagttta 600  
 tttggtgcac ctgtgcaccc agctagtatt ggttcacag tcgttccaat tcttattatg 660  
 gtttggttga tgaaatatat tgaaaaaata attgctaaat taacactagc tattactaag 720  
 agttttttgc aacctacgct agtattatta gtatcaagct gtattgcctt agttgtagtc 780  
 ggacctattg gagtaattgt tggatgaagga ttatcaaatc tagttgggca aatgtatggt 840  
 gtagctggat ggcttacatt agctatttct ggtgctatta tgccatttat tgttatgact 900  
 ggaatgcatt gggcttttgc acctattttt ttggcggcat ctattgtcac tccagacgta 960  
 ttaattcttc cagcaatggt agggctcaaac ttatgctcaag gggctgcttc gatggctgtt 1020  
 gcattaaaga gtaaaaaata taatacaaaa caaattgctt ttgcagcagg ttctctagcc 1080  
 ttacttgtag ggattaccga acctgcatta tatggtgtga ctttaaaaaa taaaaaacgg 1140  
 ctttatgcag ctatgatttg tgggtgatta gcgggattat ttgcaggtct tactatgttt 1200  
 aaagcatatc tatttctgtt ccatcttttg atagcgttgc ctcaatttat ttattctgat 1260  
 gtgccatcaa atattgttaa tgctttaatt gtggcggtca ttccggttgt tattaccttt 1320  
 gtattagctt atatatattg aatcgatgaa gaagagagtt ctagcaattt agaagttgaa 1380

```

gctggagttt caaataaaaa aatgatattt tctcctatat caggagaaat cattccgtta 1440
agcgatgtcc aggataaaac attttcagat aaactaatg gagacggagt agcgattatc 1500
ccaagtgaag gtaagggtta tgcaccattt gatgggaaaa ttacaaatat ttttccgact 1560
aagcacgcaa ttggattgaa gagtgatgag ggtgttgagt tactaattca tattggatta 1620
gatactgttg agctaaaagg tcaaggtttt attagtcatt tagaagaagg agacagagtt 1680
ttcaaaaatc agttgatttt tgaatggac ttgaatttaa tcaagactaa aggctacgaa 1740
acagttacac cagtaattgt aacgaatacc aatgattttc tagatgtatt agtattacct 1800
aataatcaga caatcgagca ttctaaggaa ttactggtaa tattataa 1848

```

```

<210> 39
<211> 246
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 39
atagctggca agtgggcaat ggtcggaaac gccaaatcat ttgggataaa gtcagccaaa 60
cgaaccgtcg tttccttgat attaaattta ttattgctgg cagttacact gataaaatgg 120
ggagccaact cctgcatttc ctgcaaggct gaaataatgt tatcattacc caggctgggg 180
tttgagggga acacttcaaa tgagagtgcg ggtgtttggc gtgacatatg taataacctt 240
ttctag 246

```

```

<210> 40
<211> 669
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 40
gaggttacta tggaatctat tttagaagtt ttaacccag ataacctagt ctttatcttt 60
aaaggatttg gcttgaccct ctatatcttc ctgattgcca tcatcctctc tactatcctc 120
ggtacgggtc tagctgtcac gagaatggc aaaaatcctg tcttacgcac tatttccagt 180
atttatatcg agtttggcgc caacgttccc aacctctctc ggatttttac tatctttttg 240
gtgttcaaaa tgaaatccac accagcaggc attacagcct ttactctctt tacatcagca 300
gccttggtcg agattattcg aggcgtgttc aatgccgtag acaagggaca gtacgaagca 360
ggaatgtcac aaggcttcac ctacgccaa atcctctact acatcattct cccacaagcc 420
atccgcaaaa tgctaccagc catcatttct cagtttggtta ccgtgattaa ggataccagt 480

```

```

ctctctact ctgttatcgc cctacaagaa ctctttggag ccagccaaat tctcatgggc 540
cgttatttcg aacagagaga ggtcttcagt ctttacatcc tgattgcct catctacttc 600
agctttaacc tagcaatttc tagcctgtct catatgctag ccaaactgtg gcaacaagct 660
gcagaataa 669

```

```

<210> 41
<211> 768
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 41
agatctctca tggctttagt agaatttaa aacgtcgaaa aatattacgg agactaccac 60
gcattccgca acatcaatct cgtttttgaa aaaggacaag ttgttgctct gcttggacct 120
tctggctctg ggaagtcac tcttatccgt acgatcaatg gtttagagac tgttgacaaa 180
ggaagtctcc tagtcaatgg gcaccaagtt gctggtgcca gccagaaaga tttgtacct 240
cttcgcaagg aagtcggcat ggtttttcaa cattttaacc ttatccaca caaagctgtg 300
ttagaaaacg taacgcttgc acccattgaa gttctaggaa ttgataaaaa agaagctgaa 360
aaaacogccc aaaaatatct ggaatttgta aatatgtggg acaagaaaga ttcctatccc 420
gccatgctat ctggttgaca aaaacagcgg atcgccatcg ctcggtgtct tgctatgcat 480
cgggaactcc tctcttttga tgaaccaaca tctgctcttg atcctgagac tatcgagat 540
gttctagcag ttatgcagaa actggcgcat gatgggatga acatgatcat cgttaccac 600
gaaatgggct ttgctcgaga ggttgccgac cgcattatct ttatggcga cggagaagtt 660
ttagtagata cgacagatgt cgataacttt ttgacaatc caagcgaacc tcgtgccaa 720
caattctcca gcaaaattat caaccacgaa agtgacaaag tcaataa 768

```

```

<210> 42
<211> 1224
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 42
gaaatgtacc gttatcaaat tggcattccc acattagaat atgatcagtt tgtcaaagaa 60
catgaattag ccaatgtatt acaagtagt gcttgggagg aagttaagtc taattggcaa 120
catgagaagt ttggtgttta caggaagaa aaattactgg cgacagctag tattttgatt 180

```

```

agaactcttc cgctaggcta taaaatgttt tacatcccaa gaggacctat attggattat 240
ggggataaag aactcttgaa ttttgcatt cagtctatta agtcctatgc tcgcagtaag 300
agagcggttt ttgtgacttt tgaccaagt atttgcctat ctcaaagttt aatcaatcag 360
gaaaagacag aatttctcga aaatctggct attattgata gtttgcaaca aatgggagta 420
aggtggtcag gaaaaacyga ggaatggga gacaccattc aacctcgatc tcaggcgaaa 480
atatacaagg aaaattttga agaagataaa ctttccaagt caacaaaaca ggctattcga 540
acagcacgaa acaaggggct tgagattcaa tatggtggac tggaactatt agattcattt 600
tcggagttga tgaaaaaaac tgagaagcga aaagagattc atttgaggaa tgaagcctat 660
tataaaaaat tgttagataa ttttaaggac aaggcctata tcaccttggc caccttggat 720
gtttctaaac gttcgcaaga gttagaagaa cagttagcga aaaatagagc cttggagag 780
acctttactg agtcgactcg aactcaaaa gtagaagcgc agaagaagga aaaagacgt 840
ttgtagagg aattgacctt cttgcaggaa tatatagatg taggtcaagc gagagttcct 900
ttagcggtta ctttgagttt ggaatttggt actacctctg tcaatatata tgctggatatg 960
gatgatgatt ttaaacgtta caatgcacca attttaactt ggtatgaaac ggctcgctat 1020
gcctttgaac gaggtatgat ctggcaaaat ttggtgggtg ttgaaaactc tctcaatggt 1080
ggactttatc attttaagga aaaatttaat ccaacgattg aagaatactt ggggtgaattt 1140
acaatgcccc ctcatctctt ctatctcttg ttaagacttg ctcttgattt ccgtaaaaca 1200
ttaagaaaaa aacatagaaa gtaa 1224

```

```

<210> 43
<211> 636
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 43
tgcttttttc agactcctaa tcgtgggtata ctaggctcagt attttataaa tatgaaggag 60
atttttatgg ctaaaaaagg taccctaaca ggtttgctcc tgtttggaat attttttggt 120
gcggggaact tgattttttcc gccttctcta ggtgctctat ctggagaaca ttttcttctt 180
gccatcgtag gttttgtctt ttacggcggtt ggtatcgccg tcttgacctt tattatttga 240
acgctaaatc ctaaaagata tatctacgag atttcaacga agatagcgcc ttggttttgcg 300
actctttacc tctcagttct ttactgttca atcggtccat tctttgtctac cccacgtact 360

```



gctacaacag cttacgaagt agggattagc ccccttttgt cggatgcaaa taaaggactt 420  
 ggcttgattg tatttacggt tctgtatttt gcgcagcct atttgatttc gcttaatcca 480  
 tcaaaaatct tagaccgcat tggacgtatt ttaacgccag tctttgcaat ttgtattggt 540  
 atcttggtcg ttctgggagc tatcaaatat ggtggaacaa gtctcaagc tgcttcactg 600  
 cttatcaagc ttctgccttt ggtacaggtt tcttag 636

<210> 44  
 <211> 2049  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 44  
 tccatcaaaa atcttagacc gcattggacg tattttaacg ccagtccttg caattttgat 60  
 tgttatcttg gtcgttctgg gagctatcaa atatggtgga acaagtcctc aagctgcttc 120  
 actgcttacc aagcttctgc ctttggtaga ggttcctag aaggttacia taccttggac 180  
 gcccttgctt cagtgccctt tagcgttaac gcagttcaaa ccttgaaaca acttgagatt 240  
 tcaagtaaga aagaatacat ttcaactatt tgggttggtg gtatcgttgt tgcccttgcc 300  
 ttacgcgtc tttacatcgg tttaggtttt cttggaaac atttccagc accagctgaa 360  
 gcgatgaagg gtggaacacc aggtgtttac atcttgtcac aagccactca agaaatcttt 420  
 ggctcaacag ctcaactctt ccttgcagct atggttaccg taacctgctt cacaacgact 480  
 gttggtttga ttgtgtcaac agctgagttc tttaatgagc gcttcccaca aatcagctac 540  
 aaggtttatg cgacagcctt taccttgatt ggatttgcta ttgccaattt gggctctgat 600  
 gcgattatca agtactcaat tccagtactg gttatcttgt accaatcac gattgctatc 660  
 gttatgattg tcaattgtcaa caaatttggt gccctttcaa aaccaggat gcagttgaca 720  
 attgctgtgg ttacagttat tgccattgca agcgtagtag gaagctcggt aaggttgagt 780  
 ttcttgcaaa tcttggttagc gttcttcctt ttgccaaggc atctctccca tggttggtgc 840  
 cagccattgt tggaaacttg ctctcattgg ttctacaaaa caagcaagaa agcgatgttt 900  
 ttgaaatgga ataatacactt aaatcacttt tgtagccaag tctacaggag tgattttctt 960  
 tttttatccg atgataaatg tgttataata ggtagcgaaa gagtggaaga aatgaatcaa 1020  
 acagtagaat atatcaagaa actgacagcc attgcgtgc caacaggctt tactcgtgag 1080  
 attgcggact atttagtcaa gactctagaa ggttttggtt accagccggt tcgcacatcc 1140

```

aagggcgggtg tcaatgtaac tattaagggt caaaatgatg agcaacatcg ctatgtgact 1200
gcccatgtag atacgcttgg tgctattgtc cgtgctgtca aaccagacgg cegtctcaaa 1260
atggaccgta tcgggtggctt tccttggaac atgattgaag gagaaaactg taccattcat 1320
gtggctagca caggtgaaaa agtatcagga accatcctca tccaccaaac ttcttggccat 1380
gtctataagg atgcaggaac tgcagaacgc acgcaagaca atatggaagt gcgtttggac 1440
gccaaagtaa ctagtgaaaa agaaactcgt gctcttggca ttgaggtcgg tgattttatc 1500
agttttgacc cacgaactgt cgtgacagag acaggtttta tcaagtctcg ccatttggat 1560
gacaaggtea gtgcggcgat tttgctcaat ctcttcgca ttataagga agagaagatt 1620
gaattgcccg taacaactca ttttgctttt tcagttcttg aagaagtggg acacggtgca 1680
aactctaaca ttctgctca ggtagtagaa tatctggctg tggatatggg agccatggga 1740
gatgaccagc aaacagacga atatacagtg tctatctgtg tcaaggatgc ttctggacct 1800
tatcactatg acttcgctca acatttgggt gctttggcga aagagcaaga tattccattt 1860
aagctggata tctatccatt ttatgggtcg gacgcttcag cggctatgtc tgcaggggca 1920
gaagtcaaac acgcctctct cgggtgctggt atagagtcta gccattccta tgagcgtacc 1980
catattgact cgggtgatcg aacagaacga atggctgatg cttatcttaa gagcacgttg 2040
gtggactaa 2049

```

```

<210> 45
<211> 1032
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 45
aaacacaatg ttgctattcc ttacgatagg gagatagata tggcaatgat agaagtgaa 60
catcttcaga aaaattttgt gaagactggt aaggaaccgg gcttgaaggg ggctttgcgc 120
tcctttattc atcctgaaaa gcagaccttt gaagcgggta aggatttgac ctttgagggt 180
ccaaaagggc agatttttagg atttatcggg gcaaatgggt ctgggaagtc gacaaccatt 240
aaaaatgctg caggaatttt gaaaccaaca tctggttttt gtoggattaa cggaagatt 300
cccaggaca atcggcaaga ttatgtcaaa gatattggcg tagtctttgg acaacgcacc 360
cagctatggt gggatttggc tctgcaagag acctacactg tcttaaaga gatttatgat 420
gtgccagact cgctctttca taagcgtatg gactttttga atgaagtctt ggatttgaag 480

```

gactttatca aggatcccg tgcgactctt tctactgggac aacgggatcg ggcggatatt 540  
 ggcggctcct tgcctccaaa tcccaaggtt ctttttttag atgagccgac cattgggttg 600  
 gacgttttcg ttaaggataa tattcgtcgg gcaattactc agatcaatca agaggaagaa 660  
 actaccattc ttttgaccac tcacgatttg agtgatattg agcaactttg tgatcggatt 720  
 ttcatgattg acaaggggca agagattttt gatggaacgg tgagccaact caaggagacc 780  
 ttggtaaga tgaagactct ctcttttgaa ctgctaccag gtcaaagtc tctcgtctct 840  
 cactatgacg gtctgtctga tatgaccatt gatagacaag gaaacagcct caacattgaa 900  
 ttgatagatt ctgcctacca gtcagctgac attatcaagc aaacctgtc tgattttgaa 960  
 atccgcgatt tgaagatggt ggatacggat attgaggata ttatcgtcgt cttctaccga 1020  
 aaggagctct ag 1032

<210> 46  
 <211> 1509  
 <212> DNA  
 <213> *Streptococcus pneumoniae*

<400> 46  
 cattcatata acatcaaaaa gggaggaact gttatggatg caatctttga cctaatacga 60  
 aaggttttca atcccatctt agaaatgggt ggacctgtca tcatgttaat cattttgaca 120  
 gtattggctt tacttttttg agtgaaattc tccaaagcgc ttgaagggtg tatcaaaactt 180  
 gccatcgctc ttacaggatg cgggtgctatc atcgggtatgc taaacactgc ttctcagca 240  
 tcactagcaa' aattcgttga aaacactggt atccaattga gtattaccga cgttggttg 300  
 gcaccacttg ctacaatcac ttggggttct gcttgacac' tatacttctt gctcatcatg 360  
 ttgattgtca acatagtgat gctagctatg aagaaaacag atacacttga tgtogatatc 420  
 ttgatatact ggcactgtgc tatcacagggt ctcttgatta aatggtatgc tgataacaat 480  
 ggtgtgagtc aagggggttc actctttatt gctacagcag ctatcgctct tgctgggttg 540  
 ttgaaaatta tcaactctga ctgatgaaa cctacatttg atgacctctt taacgcccca 600  
 agttcatcac caatgacatc aactcacatg aactacatga tgaaccaggt tatcatggtt 660  
 ttggataaga tttttgaaaa attcttccca ggccttgata aatatgactt tgatgctgct 720  
 aaattgaaca agaaaatcgg tttctgggga tctaaattct tcacgtggtt catccttggt 780  
 atcgttatcg gtattatggg aactccacat ccaattgcag gtgttgcaga tgcagataaa 840

tggcgctcttg ttatcaaagg atggttgtct cttggtttga ctgccgggtgt atctttggaa 900  
 ctcttctcac ttatcggttc atggttcac gcagccgtag aaccactatc acaaggattt 960  
 acaaacgttg ctactaaacg tcttcaagga cgtaaattca atatcggtct tgactggcca 1020  
 ttcatcgctg gtcgtgctga aatctgggct tgtgccaacg tacttgccacc aatcatgttg 1080  
 atgaagcag tgcttcttct aaaagttgga aatggtatct tgccacttgc aggtatcact 1140  
 gctatgggtg ttaactccagc tctcttggtt gtaactcgtg gtaaatgtct ccgtatgatt 1200  
 atcttcggaa cactcttgtt gccactcttc cttctttcag gtacacttat tgcaccattt 1260  
 gcaacagaac ttgctaaagg ttaggtgcc tcccagaag gtgtgagcca aactcaattg 1320  
 attactcact ctactcttga aggaccaatc gaaaaacttc ttggttgac aattggtaac 1380  
 actacaactg gtgatatcaa agcaatcctt ggtgcagtag tcttcttctg attctatatc 1440  
 ggtatctttg cttggtacag aaaaacaatg atcaaacgta acgaagagta cgcagcaaaa 1500  
 gcaaaataa 1509

<210> 47  
 <211> 366  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 47  
 tacaatatgg gtatgatatt aatgaaatta gcactatatt tattattgat actgacctta 60  
 gtcgtctgca ttactctaac caaacttttt agattaaaaa aactaggacg aaactttgct 120  
 gatttggctt ttccagtctt ggtattttgag tattacttga ttacagctaa aacctttacc 180  
 cataatttcc tccctagact ggggctagcc ctctcgatcc tagccattat tctcgtcttt 240  
 ttcttctttt tgaaaaaacg cagcttttac tacctaaat ttatcaaatt cttctggcgt 300  
 gcaggattct tattaacctt tatcatgtat atagaatga ttgttgaatt gttcttaatt 360  
 aatag 366

<210> 48  
 <211> 729  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 48  
 aatgatagga ggaactttat ggggtcatatt ttcttttttc taagtgtctt ttggcaggg 60  
 attctatcct tcttttctcc ttgtatctta cctttgttac cggctctatac aggagtgtta 120

```

ctagatgata aggatgggtc tcaggcttct agcggcaaat ttccaatctc agttactagt 180
ttattacgaa ctctggcctt tatagcagga atttcttta tttttatgtt gttgggctat 240
ggagctgggt ttttaggcga tttgctttat gcttcttggg tccaatatct tactggggca 300
attattatcc ttcttgggtt gcaccaaag gagattctac actttaaggg gctttataag 360
gaaaagaggc tacaactgca aggacagggg caaaatggta agggctatag tcaggcattt 420
ttattgggct tgacctttag ttttgccttg acgccttgcg tggggccggt tctggggctt 480
gttttggcct tggcggcttc aggtgggtca ggagcttggc agggagctgg tctcatgttg 540
gtgtatacgc tgggcttggc gctaccattc ttgcttctag ctctgacctc tagttatgtt 600
ttgaacattt tccgaaaact tcatccctat ctccgaatcc tcaaaaaagt ggggtgtttt 660
ctcattattg tgatggggctt cttgggtctg tttggaaatg cttcaatttt aagtcatta 720
tttgaataa 729

```

```

<210> 49
<211> 303
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 49
ttttggacga ctagccagtg ccgttacatg ggcattgacca atctctctca aaataggggc 60
aatcggaacc tgaacatgct tgacatgcat gccaatgca gtgtctcga tatccaatcc 120
agcatgagcc ttgataaatt caacctcaac tggatcctgc ataaacttaa aggtgcgcaa 180
ctgccccgaa cctcctgcat gaagagtagy atggacactg acaatttcca gaccaaacgt 240
ctctgccacc tgacgttcaa caacgagagc ccgattgaca tgctcacaac cttgaactgc 300
taa 303

```

```

<210> 50
<211> 1014
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 50
ttatgggaga aagaatgaa taaactgcta ttttcaaaaa tgagtctggt gacgttgcca 60
atatttagct tgttttcaca atcagttttg gcggaagaaa acatccattt ttcgagctgt 120
aaggaagctt gggcgaatgg ctattcggat attcacgagg gagaacctgg ttattctgccc 180

```

aagtttagacc gtgatcatga tgggtgtggct tgcgaattga aaaatgctcc taaggggtgct 240  
 tttaaagcaa aacagtcacac ggctattcaa atcaacacaa gttcagcaac aacaagtggt 300  
 tgggttaagc aggacggcgc ttggtactac ttgatggaa atggaaatct agtgaaaaat 360  
 gcatggcagg gaagctatta cctgaaagct gatggtaaaa tggcacagag tgaatggatt 420  
 tatgactctt cttatcaagc ttggtattat ttgaaatcag atggttctta tgcaaaaaat 480  
 gcatggcaag gagcttatta ccttaaatca aacggtaaaa tggcacaagg tgaatgggtt 540  
 tatgattctt cttaccaagc atggtattac ttgaaatcag atggttcata tgctcgcaat 600  
 gcatggcaag gaaactacta ttgaaatca gatggtaaaa tggctaaagg tgaatgggtt 660  
 tatgatgcc cctatcaagc ttggtattat ttgacatcag atggttctta tgcttacagt 720  
 acatggcaag gaaattacta tctaaaatcg gatggtaaaa tggctgtcaa tgaatgggtt 780  
 gatggtggac gttattatgt tggcgctgac ggagtttgga aggaagtcca agcaagtaca 840  
 gttcttctta gtaatgatag caatagttaa tattctgctg ctttaggaaa ggcaaaaagt 900  
 tataattcgt tattccacat gtcaaaaaaa cgtatgtata gacaattaac ttctgatttt 960  
 gataaatttt caaatgatgc agctcaatat gccattgatc atttagatga ttaa 1014

<210> 51

<211> 1239

<212> DNA

<213> Streptococcus pneumoniae

<400> 51

atgattgaaa cggagaaaaa agaggagcga gtcctgctga ttggtgtgga attgcaggggt 60  
 atggacagtt ttgacctctc catggaaaga ttggctagtt tagcgaaaac ggcaggggca 120  
 gtcgtttag atagctacag acaaaaacgt gaaaaatat attccaagac cttcgtcggc 180  
 tctggaagt tggagagat tgcgcttatg gtggatgcag aagaaatcac tactgtcatc 240  
 gtcaacaatc gtctgacccc aaggcagaat gtcaatctag aggaagtctc cggtgttaag 300  
 gtcatlgacc gtatgcagtt gattttggat atctttgcc tgcgggctcg aagccatgaa 360  
 gggagctcc aagtcacact agcccaactc aaataccttt tgctcgtctt ggttggtcag 420  
 gggattatgc tcagccgtca ggcaggggga attggttccc gtggctcctg tgaagccag 480  
 ctggagctga accgtcgtag cgttcgcaat caaatcacgg atatcgagcg ccagctcaag 540  
 gtggttgaga aaatcgtgc gactgtcaga gaaaaacgtt tggagtctag cacttttaag 600

```

attggtttga ttggttatac taatgctggg aaatcaacta tcatgaacat ctgaccagt      660
aagacccagt atgaagcaga tgagctcttt gcgactctgg atgcgacaac caagagtatt      720
catctgggag gcaatctcca agtaactttg acagataccg ttggtcttat ccaagatttg      780
ccgacagagt tgggtgcccag tttcaagtca accttggaag aaagcaagca tgtggacctt      840
ctggttcattg ttatcgatgc tagcaatcct taccacgagg agcatgaaaa aacggttctc      900
tccatcatga aagacctgga catggaagat attcctcact tgacgcttta taataaagcg      960
gatttgggtgg aggatttcac gcctacccaa acgcatata cctcatttc tgccaagtct     1020
gaggacagtc gtgaaaactt gcaagcatta ttgctagata agattaagga aatttttgaa     1080
gcatttacc tgcgagtgc tttttcaaag tctacaaga ttcatgattt agagagtggt     1140
gcaattcttg aagaacgtga ttatcaggaa gacggcgaag tgattacagg ctacatttcc     1200
gagaaaaata aatggagggtt agaagaattt tatgactga                             1239

```

```

<210> 52
<211> 267
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 52
aaagagagaa agatggtcta tttagtctta ggaattttac tgctcctact ctatgtattt      60
gcgacaccag aaagcattaa agggactgtc aatatcgtcg ctatggtatg tattttagtg     120
gcactcttga ttttattggt tctatctttt ctgaaaattt ttcaattacc aacagaaata     180
ttcctagcaa tagccatggt gatcctagct tacttttagtg ttagagacat cacactcatg     240
ccagtcaaaa aaagttaaag aagataa                                           267

```

```

<210> 53
<211> 810
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 53
actataaatg aacaaatttt aatttcggat gagatagata ttgatagtag atattctaga      60
actaaagggt actatcggtt attttataat gaagagtata ataaaataca gaataaaaaca     120
gtattagtat taggagcagg agtccttagga tggttatatat ctctaagtct aagtatgtat     180
ggagtgagga aacttattgt cgctgattac gatataatag aaccatcaa ttttaaatagg     240
caaatctttt atacagagtc ggatgttggt aaggagaaga ttaatgttct tcttgaaaaa     300

```

atacacaagt ataattcaga tgttcaggta gtacctatct ctattaaagt ttcttcagta 360  
 gaagaatttag aaaaaattgt tgcgggaatat gggagtatag attttatcgt taaagcaatt 420  
 gatacgccca ttgatattat aaaaattgtc aatcaatttg ctgtatcgca taagatatcc 480  
 tacatatcag gaggggttaa tggatgctat ctctattatg ataatatata tatccctacc 540  
 atcgggttctt gctttggttg tcgggaatata aacaaagata taaataagta cactttatct 600  
 gataagacaa agtggcgagc tacaccagag atgcctgcta ttttgggagg gataatgact 660  
 aatttaataa ttaaaatatt tctgggatgt tataatgaaa toctaataga taacgcttac 720  
 gtttataata tgagaaatca tgcctaatg caagaaaaat atgttctgga aaacggagaa 780  
 tgtccaatct gtaaaaaaat aataaagtga 810

<210> 54  
 <211> 393  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 54  
 aaaaataata aagtgaaga taacaatatt agagcgaaaa catttatctg ttcagtttgt 60  
 ttttctctat tatcaggagg agtagctttt ttatctgcta ttgggcagtt caactgttata 120  
 gaaacacaat taatagtatt gttcttgggt attatttttg ctatatatta tgcttactac 180  
 aataaaaata ttcaaacatc attggaaaat atagtatggc ttttttcac gtttgagatt 240  
 ttatttttgc ttgttaattt tagaacattt attcagttac cagtggatatt ttttatttgt 300  
 atgataatat ttttaattgt gtggatattt attatgttag gtatagtgtg tcttagttat 360  
 tatataactt tattatttag caaggaggct tag 393

<210> 55  
 <211> 750  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 55  
 atttttggat ttttagaacc atctggttct ggaagacca caacgattaa tattctgact 60  
 gggcagttcc ttgccgataa aggacaatct attatttttg gacaaaaatc tcaaaattta 120  
 acaagcgggt aattaaagag aattggattg gttagcgata caagtggatt ttatgagaaa 180  
 atgtctctgt ataacaatct tcttttttat agtaaatatt ataatttag taaatcacgt 240



gttgataatt tgttaaacg agtaggatta tatgatagtc gcaagatggt agcaggaaaa 300  
 ttatccactg gaatgaggca acgaatgott ttagcacgag ctcttatcaa caaccccgt 360  
 gtactctttc tggatgaacc gacctcaggt ctgatgccca caacttctcg aacaattcat 420  
 gagttaattt tagaattgaa aacagcaggg acaacgattt ttctaacgac tcatgatatg 480  
 aatgaagcaa ctcttttatg tgattatggt gccttattaa ataaagggaa attagttag 540  
 caaggagctc cttctgaact cattcaaaga tataataaag ataaaaagat taagggtaca 600  
 gattataatg ggaatcagat aacttttgat ttacatcac tagaacaggt atctcagact 660  
 gatctggaaa atatttttct aattcattca tgtgagccta ctttagaaga tattttttatc 720  
 acattaacag gaggaaagct aaatgcttaa 750

<210> 56  
 <211> 777  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 56  
 ggaggagtaa ggaatgaatt tttcatttgt aatctgttac gagtcgttca atcacctcga 60  
 ttttatatgt ctttattttt gaccttctt tgcattgagt taggaaattt ccttgcttcc 120  
 aatggtatct ataaaaattg aggtttatcg attttttttg ccgcttcttc tattcgagga 180  
 ttttcacoga ttagcctagt agctgcactt atctgtacac tgcctatctc tagtcagata 240  
 atagaggatg ctgagagtca ttttctaaca gcacaattgt gtcgaatttc taaaaagaag 300  
 tatctggcta ttgtgggtag tactgttaatt atttcttctt ttctagtctt tttctctccc 360  
 tatttattat tattaggaat taatctttta gtgactcctt atcaggaaat ttatattgga 420  
 gattatagtg gtgccttaaa agaattattt gattccaatc agtttctcta tagtcttgta 480  
 acgactctct ggtatggagt ttggggcgct gtgttctcta tttttggact agctagtgtc 540  
 ttgctagtga agaaaaaat aggagctatt ttcattccag ttgcctatat gatggtgggt 600  
 ggtatttttt gggctatttt agggctatct tacttagaac ctgtgacaac gctagctttg 660  
 ggatatacaga aagatatcag tctttcctta gttagtgtc atcttgcttt tattttattt 720  
 gttagtgtgt ttggtgttta tggtagatct tttctacatt cagaggacta tgtataa 777

<210> 57  
 <211> 777  
 <212> DNA

<213> *Streptococcus pneumoniae*

<400> 57  
 ttatctatatt tatattttga atcgagggaa tggagatata ggacgagatt tattttctca 60  
 gttgctatatt ttagtctttc tttttcaag cattttctat ttactcgtc aaaagaggag 120  
 gtttatagaa tgaatgaaat aattacatta aaaaatatg agttgaaatt aaaaaagaca 180  
 tgtgtttttc aaacacctaa ctttagttgt aaacaggggg aaattatagg aattactggt 240  
 gcgaatggct cagggaaaag tgtattgttt aaattaatag ctggtttata tagtccgtct 300  
 tatggagaag tgttaatcaa tggggaaaat attgttctg agagaaaaat tccagctaatt 360  
 ttgggagcct tgattgaaga acctgggttt ataaattatt atagtggcct taagaattta 420  
 caatatttgg caagcatatg aggagtagtt ggtaatcagg aaatcaatga tacactgaaa 480  
 atagtgtgtc tatatgagca aaaagaccag aaagttaaaa cttattcgtc aggtatgagg 540  
 aaaaagctag ggattgtctc agcaattatg gagaatccct ctattctttt actagatgaa 600  
 cctatgaatg ccttggtata atcaagtgtg gaaaatatga gaacattgtt tagaaagctc 660  
 tctagtgaat aagggaacac aattttgatt gctagtata gtgaagagga tattcgtatc 720  
 ttatgtgata aagtatatgc aatagaagat aaagtatgta cactgtgttc agattga 777

<210> 58

<211> 759

<212> DNA

<213> *Streptococcus pneumoniae*

<400> 58  
 atgtctgaaa ctatcttaga aatcaaggaa ctaaaaaat ccttcggaga caatcccatc 60  
 ctccaaggac tttctctaga aatcaaaaaa ggggaagttg ttgtcatcct agggccatct 120  
 ggttgtggga aaagtaccct ccttcgttgc ctcaacggct tagaaagtat tcaaggtgga 180  
 gatattcttc tggatgtgtc gtctatcgtt gaaaataaaa aagattttca cctagttcgc 240  
 caaaagattg gcatggtctt tcaaagttat gaactctttc cccatctgga tgtcttaca 300  
 aaactcatcc taggccttat caaagctcaa ggaagggaca agaaagaagt aacggaagaa 360  
 gctttgcaat tactagagcg tgtcggtttg ctggataaac aacatagcct tgcccgtcaa 420  
 ttatctgggt gacagaagca acgtgttgca attgtccgtg ccctcctaatt gcatccagaa 480  
 atcatccttt ttgacgaggt gactgcttcg ctggatccag aaatggtgcg tgagggtgctg 540  
 gaacttatca atgatttggc ccaagaaggc cgtaccatga ttttagtaac ccacgaatag 600

cagtttggccc aagccattac tgaccggatt atcttctctcg accaagggaa aatcgctgaa 660  
 gaaggaacag ctcaagcctt ctttaaccaat cgcgaacca aacgagccca ggaattttta 720  
 aacgtctttg acttttagcca attcgggtca tatctataa 759

<210> 59  
 <211> 672  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 59  
 tccattgttg aacaatatct accactatat caaaaggcat tctttctgac cttgcatatt 60  
 gcagtttggg gaatttttggg atcctttctcg ctcggttttaa tcgtagtat catcgacat 120  
 tatcgaatcc ttgttttggc gcaagtagcg acagcctaca ttgaattgtc acgtaatacg 180  
 ccccttttga ttcaactctt ctttctctac ttcggtcttc ccgaatcgg gattgtctcta 240  
 tcttcagaag tctgtgcaac gcttgggctt gtcttttttag gaggctccta tatggcagaa 300  
 tctttccgaa gtgggttggg agccatcagt caaaccagc aggagattgg cctcgtatt 360  
 ggtctgacac ctctacaggt cttttactat gtggttcttc cgcaagcaac agcgggtggca 420  
 ctccctctct ttagtgccaa tgtcattttc cttatcaagg aaacctctgt tttctcagca 480  
 gtggcttttg cgcacctcat gtacgtgcc aaggatttga ttggtctcta ctatgagaca 540  
 gacattgocg tagctatgtt ggtagttgct tatctaata tgctgtacc catctcactg 600  
 gtcttttagc gtagagaaag gaggctcgc catgcaggat tcgggaatcc aagtactctt 660  
 tcaaggaaat aa 672

<210> 60  
 <211> 1386  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 60  
 atgggtctgg aactacgagc gattcagtc ccaatcttct ctgagcgtt tgattttact 60  
 ttctatgcgc aagcctttac cttgttagtt gggagtagtg ggtctggaaa atccagcctc 120  
 ttccaatga ttgccaaagt tagttctctt cccatagcgt gtcaagtcct gatagatggg 180  
 agcgaggta gtcagcttct tatcgtcga cgtgtccaga cggttggtat tctcttgcaa 240  
 aatcctaata atcaatttac catggagagc ttgtttgagg agttgggttt taccatggaa 300

```

aatatcggct atcaccttca ggaaattgat tctaaaatag cagaggttgt ccagcaatgt 360
cgttgcaagg acatcttgca ccgtctcacc catcacttat caggtgggga aaagcaaaaa 420
gcagcgctgg ctgtctctct tgccatgaat cctaggtctc atctcttgga tgagcccttc 480
gcttccattg accgcaagag cagaatcgag atattggaga ttctaaaaga gttgggtctat 540
gatgggaaga cagttatctt gtgcgaccat gatttatctg actataaagc ctatatcgac 600
catatgggtg agctaagaga cggaaaacta aggggaagtgt ttcaaatccc ttcctatgag 660
atgacacagg ttgcttcaaa ggaagttgct tctagcccg aactattcca tatgaaccgt 720
gtgactggtg agcttggtaa tgcacctc ttttcaattg ctgatttcac attctatcaa 780
gggatttctc gtatcctggg tgacaatggt gtcgggaaat caaccctctt tcgggtctatt 840
cttcaatttc aaaagtataa ggggagcatt acttggaagg gttcggctct gaaaaagaaa 900
aagagtttgt atcgtgatct gactggtgtt gttcaggaag ctgagaagca gtttatccga 960
gtcagctctg gagaggagct tcaattagat ggacctgatt ctgaaagaaa tcagcggatt 1020
tttcaagctt tacgatattt tgatttgag caggcagtcg ataagagtcc ctatcaatta 1080
agtgggtgct agcaaaaaat tcttcagctc ctgacctatc tgaccagtaa ggcctccctg 1140
atcttgctag atgaaccttt tgcaggtttg gatgatagag cctgccatta tttttgcaag 1200
tggtattgtg aggagaggaa tcaaggaaga agttttctgc tcattagtca tcgttttagac 1260
cctttgattt ctgtggttga ttattggatt gagatgacta gtcagggctc tcggcatgtg 1320
aaagaagtga ccattaccaa accacttaca tctcagagta gcaataccca agggggaggtg 1380
agatag 1386

```

```

<210> 61
<211> 1212
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 61
ccatcatctc ttgttttttt gtggtacaat agagctatga aacattttga tactattgtc 60
atcggtgggg gacctgctgg tatgatggct acgatttcca gtagctttta tggacagaaa 120
accctcctca tcgaaaaaaa tcggaaactt ggaaaaaat tagctgggac tgggtgggga 180
cgttgcaatg tgaccaacaa tggtagctta gacaacctgc tagctggaat tcttggaac 240
ggacgcttcc ttacagtggt tttctcccag ttcgataatc atgacatcat caacttttt 300

```

```

acagaaaatg gtgttaaact taaggtcgaa gaccacggac gcgtctttcc agccagtgac 360
aagtctcgga ctattatcga agctttggaa aagaaaatca ccgaactagg tggccaagtt 420
gtactactaaa tagaaatcgt ttctgttaaa aaagtagatg accagtttgt ccttaagtca 480
gcggtatcaaa ccttcacttg tgagaaactc attgtcaca caggtggtaa gtcttatcct 540
tcgactgggtt cgactgggtt ttgtcacgag attgctcgcc attttaagca taccatcacc 600
gatcttgagg ctgctgaaag tcctttatta acagattttc cacataaagc cttacaaggt 660
atttctcttg acgatgtgac cctaagttat ggtaagcatg tcactactca tgatttactc 720
tttaccactc ttggtttgtc aggtcctgct gccctacgca tgtctagctt tgtcaaagg 780
ggggagggtc tctcactcga tgttttgctt caactttctg agaaggactt ggttacattt 840
ctagaagaaa atcgggaaaa atccttgaaa aacgctttaa aaaccttgtt accagaacgc 900
ttggccgaat tttttgtaca aggatatactt gaaaaagtca aacaactgac tgaaaaggaa 960
cgagaacaac ttgtccagtc cattaaagaa cttaaaatc ctgtaactgg aaaaatgtcc 1020
cttgcaaagt cctttgttac caagggtgga gtcagtcctc aggaatcaa tcctaaaacc 1080
cttgaagta agctggtaac tggcctccac ttgtcaggcg aagttatgga tatcaatgcc 1140
cacacgggtg gctttaaact cacttctgcc ctctgtaccg gctgggtggc gggaagtctg 1200
cattatgatt aa 1212

```

```

<210> 62
<211> 264
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 62
ggagacaaaa agatgaagaa aaaatttgcc ctatcgtttg tggcgcttgc aagtgtagca 60
cttcttgcag cctgtggaga agtgaagtct ggagcagtea aactgcttg taactcagta 120
gagggaaaaga caattaaat cgggtttaac ttgaagaat caggttcttt agctgcatac 180
ggaacagctg acaaaaaagg tgcccaattg gctgttgatg aaatcaatgc cgcagtggtgta 240
tcgatggaaa acaaatcgaa gtag 264

```

```

<210> 63
<211> 783
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 63
gaaggaggaa caaaactaat ggcattactt gaagtaaac agttaaccaa acatttttgg 60
ggtctaacag ctgttggaga tgtgactctt gaattgaacg aaggggaact ggttggatta 120
atcggtccaa acggagctgg gaaaaccacc cttttcaacc ttttgacggg tgtttatgaa 180
ccaagcgagg gaacagtaac cctagatggt caccctttga atgggaaatc accttataag 240
attgcctctt tgggactttg acgtactttc caaaatatcc gtctctttaa agatttaaca 300
gttttagata atgttttgat tgccttttga aaccatcaca aacagcatgt ttttactagt 360
ttcttacgct taccagcttt ttacaagagt gaaaaagaat taaaggctaa agctttggaa 420
ttgttgaaaa tctttgattt agatgggtgat gcagagactc ttgctaaaaa tctttcttac 480
ggacaacaac gtcgttttga aattgttcgt gcccttgcta cggaacctaa aattctcttc 540
ttagatgaac cagcagcagg tatgaacca caggaaacag ccgaattgac tgagttaatt 600
cgtcgtatca aagatgagtt taagattaca atcatgttga ttgaacacga tatgaatctg 660
gtcatggaag taacagaaacg tatctacgta cttgaatatg gccgtttaat cgctcaagga 720
actccagacg aaattaaagc caataaacgc gttatcgaag cttatctagg aggtgaagcc 780
taa 783

```

```

<210> 64
<211> 705
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 64
aaaggaactc acatgtcaat tattgaaatg agagatgtcg ttaaaaata cgacaacgga 60
acaactgtc tacgcggtgt ttgcggttagc gttcaaccgg gggaatttgc ttacatcgta 120
ggaccttcag gagcagggaa gtcaactttt attcgttctc tgtatcgtga agtaaaaaatc 180
gataaaggaa gctatcagtt tgcgtggttt aatctgggta agatcaaaaa gaaagatgtc 240
ccgcttctac gtcgtagtgt tgggggttgc ttccaggatt ataaattggt accaaagaaa 300
actgtctatg aaaatatgtc ttacgctatg gaagtaatcg gggaaaatcg ccgtaatatc 360
aaaagacgag tgatggaagt ttggaccttg gttggattga agcataaggt tcgtctcttc 420
ccaaatgaac tctcaggttg ggagcaacag cggattgcga ttgcgcgtgc aattgtaaat 480
aatcccaaag tattgatagc tgatgagcca acaggaaatc tggatccggg taattcatgg 540
gaaattatga atctcttgga acggattaac ctacaaggaa caactatttt gatggcgact 600

```

cataatagcc agattgtaaa taccttgccg caccgtgtca ttgccattga aaatggccgt 660  
 gtcgttcgtg acgaatcaaa aggagagtat ggatacgatg attag 705

<210> 65  
 <211> 2181  
 <212> DNA  
 <213> *Streptococcus pneumoniae*

<400> 65  
 atgatgaaag atacattcaa aaatgtcttg tctttcgaat ttggcaaaa attcggtaag 60  
 gctttgatgg tagttatcgc ggttatgccg gctgctgggt tgatgatttc aatcggtaag 120  
 tctatcgtga tgattaaacc aacctttgca ccacttgta tcacaggtagg aattcttgag 180  
 caaatcgggt ggggggttat cggtaacctt cacattttgt ttgccttagc cattggagga 240  
 agctgggcta aagaacgtgc tgggtgtgct ttgccgctg gtcttgccct catcttgatt 300  
 aacgtatca ctggtacaat ctttggtgta tcaggcgata tgttgaaaaa tccagatgct 360  
 atggtaacta ctttcttttg tgggtcaatc aaagttgctg attactttat cagtgttctt 420  
 gaagctccag ccttgaaact ggggggtatc gtagggatta tctcagggtt tgtaggggca 480  
 actgcttaca acaaatacta caacttcgt aaacttcgt atgcactttc attctcaac 540  
 gggaacgtt tctgaccatt ttaggtatt cttcgttcag caatcgctgc aattctactt 600  
 gctgcttctt gccagtagt tcaaacaggt atcaataact tcggtatctg gattgccaac 660  
 tcacaagaaa ctgctccaat tcttgacca ttctgtgat gtactttgga acgtttgctc 720  
 ttgccatttg gtcttcacca catgttgact atccaatga actacacagc tcttggtggt 780  
 acttatgaca ttttaactgg tgcagctaaa ggtactcaag tattcggtca agaccacta 840  
 tggcttgcgt gggtaacaga ccttgtaaac cttaaaggta ctgatgctag tcaatatcaa 900  
 cacttgtagt atacagtaca tccagctcgt ttcaaagttg gacaatgat cggttcattc 960  
 ggtatcttga tgggtgtgat tgttgctatc taccgtaatg ttgatgctga caagaacat 1020  
 aaatacaag gtatgatgat tgaacagct cttgcaacat tcttgacagg ggttactgaa 1080  
 ccaatcgaat acatgttcat gttcatcgca acacctatgt atcttggtta ctacttggtt 1140  
 caaggtgctg ccttcgctat ggctgacgtc gtaaacctac gtagtcactc attcgggtca 1200  
 atcaggttct tgaactgtac acotattgca atcagtgctg gtattggtat ggatacgtt 1260  
 aacttcgtt gggtaactgt tctctttgct gtaatcatgt actttatcgc aaacttcagt 1320

```

attcaaaaat tcaactacgc aactccaggc cgcaacggaa actacgaaac tgctgaaggt 1380
tcagaagaaa ccagcagcga agtgaaaggt gcagcaggct ctcaagctgt aaacattatc 1440
aaccttcttg gtggacgtgt aaacatcggt gatgttgatg catgtatgac tcgtcttcgt 1500
gtaactgtta aagatgcaga taaagtagga aatgcagagc aatggaaagc agaaggagct 1560
atgggtcttg tcatgaagg acaagggggt caagctatct acggtccaaa agctgcacatt 1620
ttgaaatctg atatccaaga tatccttgat tcagggtgaaa tcattcttga aactcttcca 1680
agccaaatga ctgaagcaca acaaaacact gttcacttca aagatcttac tgagggaagtt 1740
tactcagtag cagacgggtca agttgttgct ttggaacaag taaaggatcc agtatttgct 1800
caaaaatga tgggtgatgg atttgcagta gaacctgcaa atggaaacat tgtatctcca 1860
gtttcaggta ctgtgtcaag catcttccca acaaaacatg cttttggtat tgtgacggaa 1920
gcagggtctg aagtatttgt tcacatttgt ttggacacag taagtcttga aggtaaacca 1980
tttacagttc atgttgctga aggacaaaaa gttgcagcag gagatctcct tgtcacagct 2040
gacttgatg ctatccgtgc agcaggagct gaaacttcaa cagtatgtgt cttcacaaat 2100
ggtgatgcaa ttaaatcagt taagttagaa aaaacagggt ctcttgacgc taaaacagca 2160
gttgctaaag tagaattgta a 2181

```

```

<210> 66
<211> 1551
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 66
ggaattaaaa tgagtatttt agaagttaaa aatctgagtc acggttttgg tgaccgtgca 60
atttttgaag atgtgtcctt ccgtctcttc aaggaggaaac atatcgccct ggtcggtgcc 120
aatggtgaag gaaaatcaac ctttatgagt atcgtgactg gtaaatgct gccagatgaa 180
ggaaaggttg agtggtccaa atatgtgacg gctggttact tggatcagca ctctgtcctt 240
gctgaaagac agtcggtgag tgatgttctc cgtacggctt ttgatgagct tttcaaagct 300
gaagctcgta tcaatgacct ctatatgaaa atggctgaag acggcgcgga tgttgatgct 360
ctcatggaag aagtaggaga acttcaagac cgtctggaga gtcgtgattt ctataccttg 420
gatgctaaga ttgacgaagt agcgcggtgt cttggtgtta tggacttttg catggatacg 480
gatgtaactt ctttgcaggt tgggcaaaaga accaagggtgc ttttggcaaa acttctcctt 540

```



gaaaagcctg atatcttctg gttggacgag cggaccaact acttggatgc tgagcatatt 600  
 gattggctca agcgctatct ccaaaactat gagaatgcct ttgttctcat ttcgcaogat 660  
 attccattcc tcaatgacgt tattaatatt gtctatcatg tggaaaatca acagctgacg 720  
 cgttactctg gtgactacta ccagttccaa gaagtattatg ctatgaagaa atctcagcta 780  
 gaggcagcct acgaacgcca gcagaaagag attgcagacc tcaaggactt tgttggtcgt 840  
 aataagccc gtgttgcaac tctgaatatg gctatgtctc gtcaaaagaa attggataag 900  
 atggatatta tcgaactcca aagtgcagaa ccaaaaccat cctttgatatt caaaccagct 960  
 cgtacaccag ggcgccttat cttccaagcc aagaacttgc aaattggta cgaccgtcct 1020  
 ctactaagc ctttaaatct taccttcgaa cgcaatcaaa aggttgcgat tatttggctg 1080  
 aatggtattg gaaaaacaac tctcttgaag agtctcttgg gcattatctc gccaatgcct 1140  
 gggaagtgg agcgtggaga ttatttagaa cttggttatt ttgagcagga agtagaagcc 1200  
 ggtaatgcc aaactcctct tgaagctgtc tggaatgcct ttctgtccct taatcaagca 1260  
 gaagtcctg cagcccttgc ccgttgggt ttgacaacca aacatattga aagccagatt 1320  
 caagtattat cagggggaga gcaagccaag gttcgtttct gtctcttgat gaatcgtgaa 1380  
 aacaacgttt tagtgctgga cgagccgacc aaccatttgg atgtggatgc aaaggatgag 1440  
 ctcaaacgcg ctctcaagaa atataggga tctatcctta tggctcgcga cgagccagac 1500  
 tttttgaag gctggataga ccaaatatgg gattttaata atttaactta a 1551

<210> 67

<211> 822

<212> DNA

<213> Streptococcus pneumoniae

<400> 67

atttttgaga ggcagagaat gaaaaaacta gcaacccttc ttttactgtc tactgtagcc 60  
 ctactcgggt gtgacgagc ccaacgcagt ctgcgtgggt atgattatgt tgattccagt 120  
 cttgtcgtctg aagaaggttc caaagtagct gcccaatctg ccaaggagtt aaacgatgct 180  
 ttaacaaacg aaaacgcca tttccacaa ctatctaaag aagtgtctga agatgaagcc 240  
 gaagtgatgt tccacacaag ccaagtgat attgcatta aactcttccc taaactgcct 300  
 cctctagcgg ttgaaaattt cctcactcac gccaaagaag gctactataa cggtattacc 360  
 ttccaccgtg tcactgatgg ctttatggc caaactggag atccaaaagg ggacggtaca 420

```

gggtgctcagt ccactctggca tgacaaggat aagactaaag acaaaggaac tggtttcaag 480
aacgagagatta ctctcttattt gtataacatc cgtggtgctc ttgctatggc taatactggt 540
caaccaaaca ccaatggcgag ccagttcttc atcaacaaa actctacaga taccctcttc 600
aaactcccta caagcaagta tccacagaaa attattgaag cctacaaaga aggtggaaac 660
cctagtctag atggcaacaa cccagtcctt ggtcaagtga ttgacgggat ggtggtgtgt 720
gataagattg ctaaggccga aaaagatgaa aaagacaagc caactactgc tatcacaatc 780
gacagcatcg aagtgggtgaa agactacgat ttaaatctt aa 822

```

<210> 68

<211> 1368

<212> DNA

<213> *Streptococcus pneumoniae*

<400> 68

```

aagattacaa aggagttttc aatgagagaa tatgatatca ttgctatcgg tggaggtagc 60
ggaggaatcg ctaccatgaa ccgtgctggt gaacatggag ccaaagcagc cgttattgag 120
gaaaagaat taggtggaac ctgtgtcaac gtcggttgtt ttccataaaa aatcatgttg 180
tacggggcgc aaatcgctga gactttccat caatttggag aagactacgg cttaagact 240
actgatctta actttgactt tgcaacccta cgtcgcaatc gtgaagccta cattgatcgc 300
gctcgttctt cttatgatgg tagttttaaa cgcaacgggt tagacttgat tgaaggatcat 360
gctgaatttg tagattctca tactgtgaag gtaaatgggt aactgattcg tgctaacaat 420
atcgtgattg ctacagggtc ccatccaagt attcctaata ttctggtgac tgagctaggt 480
ggctcttctg atgatgtatt tgctgggaa gaacttccag agtcaattgc cattctaggc 540
gctggttata tcgccgttga attagctggc gtactccaca cttttggtgt caagacagat 600
ctctttgttc gcccgatcg tcctttacgt ggttttgatt cctacatcgt tgaagggttg 660
gtcaaggaaa tggaaagaac aaacttacca cttcacactc acaaagtccc tgtcaagtta 720
gaaaaaacta ctgacggcat taccattcat ttcaagatg gtactagtca cacagctagc 780
caagtatatc gggctacagg tcgccgtcca aacgttaagg gcttgcaact tgaaaaagct 840
ggagtgaatc tgaacgaacg tggctttatc caagtggatg aatacaaaa tactgttgtt 900
gaggaatct atgctctagg tgatgtaacg ggcagaaaag aactgactcc agtgcaatc 960
aaggccggac gtaccctatc tgaacgtctc ttaaacggaa aaactactgc aaaaatggat 1020

```

tactcaacta ttccaactgt tgtcttttca caccctgcta tcggaactgt tggtttgaca	1080
gaagagcaag ctatttaaaga atacgggtcaa gaccaaataca aggtttataa atcaagcttt	1140
gcattctatgt actctgcttg cacttgcaac cgtcaagaat ccggtttcaa actcataaca	1200
gctggttcag aagaaaaagt tgtcggactt catggaattg gctacggcgt tgatgaaatg	1260
attcagggat ttgcctgttc tatcaaatg ggagcaacca aggctgactt tgatgcaact	1320
gtggcaattc acccaactgc atctgaagaa tttgtaacca tgcgttaa	1368

&lt;210&gt; 69

&lt;211&gt; 1338

&lt;212&gt; DNA

<213> *Streptococcus pneumoniae*

&lt;400&gt; 69

aagatgttca gtaaacttaa aaaaacatgg tatcgggatg acttttagtta ttttatccgc	60
aacttcgggtg tottccacct gattttttct acaatgactc tgattatttt acaagtcagt	120
cattcgagtc tttatacttc ggtggacgat aagcttcagt gactgagtga aaatcctcaa	180
gcagttatlc agctggctat aaataggcca acagaagaga ttaaagattt agaaaaatgct	240
agggcgagc ctagtaaaagt agaaataaaa cctaattgta gttccaatac ggaagtcatt	300
ctctttgata aagactttac tcaacttctt tctggaaatc gatttttggg cttggataag	360
attaagttag aaaagaaaga actaggacat atctaccaga ttcagggttt taatagctat	420
gggcaggaag aaatctatcg tgtgattttg atggagacca atattagtgc ggtttcaacc	480
aatatcaagt atgctgctgt cttgattaat accagtcagt tggaacaggc tagtcaaaag	540
catgagcaat tgatttgtgt cgtgatggct agtttctgga ttttgtcttt acttgccagt	600
ctctatctag ctagggctcag tgttaggccc ctgcttgaga gtatgcagaa gcaacagtct	660
tttgtggaaa atgcagctca tgagttacga actccactcg cagttttgca aaatcgctta	720
gagacccttt ttcgttaagcc agaagctacc attatggatg tgagcgaaag cattgcatcg	780
agtttggaag aagtcggaat tatcggtttt ttaacgacaa gcttgctgaa cttagctcgg	840
agagatgatg ggattaaagc ggagcttgca gaagtccaa ctagcttttt taatacaact	900
ttcacaact acgagatgat tgcttcggaa aataatcgtg tcttcggttt tgaaaaatcgt	960
atccatcgaa caattgtcac agatcagctt cttctgaaac aactgatgac cattcttttc	1020
gataatgcg tcaagtatac tgaggaggat ggtgaaattg attttcttat ctcggcgacc	1080

```

gatcgcaatc tttatttact tgtttctgat aatggaatcg gtatttcgac agaagataaa 1140
aagaaaattt ttgaccgttt ttatcagta gacaaggcta gaaccggca aaaagtggt 1200
tttggtttag gattatccct agccaagcaa attgtagatg ctctaaaagg aactgttact 1260
gtcaagata ataaacccaa gggacaatc tttgaagtga agattgccat tcagacacca 1320
tctaaaaaga aaaaataa 1338

```

```

<210> 70
<211> 1092
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 70
gattgtaatt ttcttacggg catgattctc tccttaacag tacataccta ttttatcatt 60
tttcggcag agaattatta cagaaggtt acaaaaagaa taaagtccct tttcattttc 120
aaagcatggc tgatttttga gaaatgtggt ataatttttc ttatggaaaa gattgtcatt 180
acagcaactg ctgaagatga tgaacaagtt gaacaactac tcgaagctgg cgtagaccgt 240
atctatgtcg gtgagaaaga ttttggctct cgtctgccaa cgacctttag ttatgaccaa 300
ttacgtgaaa tcgctaagtt ggttcatgat gctggtaagg aattgatcgt tgcggtoaat 360
gctctcatgc accaagatat gatggaccgt atcaagcctt tcttaacett cttggaagaa 420
atcaagacag actatattac gattggggat gcaggcgtct ttacgtagt taaccgcat 480
ggttattcat ttaagaccat ctacgatgct tcaaccatgg taactagcag tcgtcagatt 540
aactctggg gacaaaaggc tggcgcatct gaggctgttt tggcgctga aattccatca 600
gtggaacttt tcaaaatgcc agagattttg gaaattcctg ctgaagtttt ggtttacggt 660
gctagcgtca tccatcatc taaacgtcca ctcttgcaa actactataa ctttacacat 720
atcgatgatg aaaagacgca taaacgtgac ctcttcttgg ctgagccaag tgatccagag 780
agccactatt ccatttttga agataatcat gggaccata tctttgcaa caatgacctt 840
gatttgatga tcaaatatc agaattgggt gagcatggct ttactcgtg gaaactagaa 900
gggctctaca ctctcggtca gaactttgtt gagattgcaa aactctttat ccaagcgcgt 960
agcttgatc aagagggcaa ctttagtcat gctcaagcct tcttgctgga tgaagaagtt 1020
cgtaaacttc accctaaaaa ccgtttcctt gatacaggat ttatgacta cgatcctgac 1080
atggttagat aa 1092

```

```

<210> 71
<211> 765
<212> DNA
<213> Streptococcus pneumoniae

<400> 71
ataaattgga taaaaaattt aaattttaat tcaaacatg ttataatcat acagtattct 60
attttagaaa gcagtgtagc tatgaatttt tcttttttac ctaagtattt acctattttt 120
aactatgggg ctgttgtgac gattcttatt tctatctgtg ttatcttttt ggggaactatt 180
ttgggtgttg tcttggtctt tgggcaacgt tcaaagtta aaccgcttgt ttggttggcc 240
aacttgtagc tttggatttt cgtgggaca cggatgatgg ttcaaattat gattgccttt 300
gctcttatgc atatcaatgc tccgactatt cagattggaa ttttaggtgt tgatttttcg 360
cgtctgattc cagggtattt gattatctct atgaatagt gtgcttatgt ttccggagact 420
gttcgtgccg gaatcaatgc ggttcacaaa ggtcagctag aagcggctta ttccgtaggg 480
attcgtccta aaaatgcgat gcgttatgtg attttggcac aagcagctca aaatatcttg 540
ccagcattgg ggaacgaatt tatcaccatt atcaaggaca gctccctctt atcagctatt 600
ggggtcacgg agttgtggaa tggggctaca acagtttcta caacaaccta tctaccttta 660
acaccacttt tatttgcagc attttactac ttgattatga cctctattct gacagtagcc 720
ttgaaagctt ttgaaaaaca tatgggacaa ggagataaga aataa 765

<210> 72
<211> 741
<212> DNA
<213> Streptococcus pneumoniae

<400> 72
gaaataatga cagaaacctt gataaaaatt gaaaatttac ataaatcctt tggaaagaat 60
gaagtattga agggcatcaa cctcgagatt aaaagaggag aagttgtcgt tatcatcggt 120
ccttcaggga gcgggaaatc taccttgctt cgtctatga atttgttgga agaagcaacc 180
aaggggaagg ttatctttga gggagtcgat attacggaca agaagaatga cctgtttgcc 240
atgcgtgaga agatgggcat ggtttttcaa caattcaatc tctttcctaa tatgactgtg 300
atggaaaata tcaccttgct ccctatcaag accaaagggt acagtaaggc cyttgcagag 360
aaaagagctc aggaactttt ggaaaaagtt ggtttgccag ataaggcaga cgcttatcca 420

```

```

cagagtttgt caggtggcca gcaacagcgg attgccatcg cgcgtgggtt ggctatggaa 480
ccagatgttt tgctctttga cgagccaact tcagccctag atcctgagat ggttggagaa 540
gttctggctg ttatgcaaga tctagccaag tcaggaatga ccatgggtat cgtaacacat 600
gagatgggat ttgcccgta ggtggcagat cgtgtcatct ttatggcaga cgggtgggtt 660
gttgaagacg gaacacctga gcagattttt gaacaaacc aaggacaaa gactaaagac 720
ttcttgagta aggttttata a 741

```

```

<210> 73
<211> 261
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 73
ttcacaactt ataggagggtg tactatgaaa atcttataaa gttacatatt ggaactctgt 60
tttattttta gttttgcttt accttttata aaaggaacca atgcagataa tggtagatgc 120
tttgtggaaa cctattacgg ttttactttt ttgatggaa atgctattgt aacagctgtc 180
tttatttgtt cgttcttaat tgctttctta ctaaaaaacg atggacgaaa tggattgctg 240
cgggtagtta ttgcttttta g 261

```

```

<210> 74
<211> 1548
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 74
aagggaagac acatggcaca cgaaaaatgtc attgagatgc gtgatattac caagggtgtt 60
ggtggatttg ttgccaacga caaaatcaac ttgcacctac gaaagggtga aattcatgca 120
cttttaggag aaaaatggggc tggtaagtcc acgctaatac acatgttagc aggccttctt 180
gaaccaacta gtggtgaaat cgcgggtcaac ggtcaagttg tcaatctcga ctccccatct 240
aaagcagcta gcttgggaat cgggatgggt caccagcact ttatgttggt tgaagccttc 300
acagtggctg aaaacatcat tttaggtagt gaattgacta aaaatgggtg gctagatatc 360
gctggagcta gcaaaagaaat caaggctctt tctgaacgtt atggcttagc tggtagacct 420
tctgccaagg tagcagatat ctcatgttga gcccaacaac gtgtagaat tttaaaaca 480
ctttatcggg gggctgatat ccttatcttt gacgaacca cggctgtttt gactccatca 540
gaaattgatg agttgatggc tattatgaaa aatcttgtca aagaaggaaa atcaattatc 600

```

ttgattacc	acaaattgga	tgaaattcga	gcagtttctg	accgtgttac	agttatccgt	660
cgtgggaat	caattgaaac	cgttgaaatt	gcaggggcta	ccaatgctga	tttggcggaa	720
atgatggtag	gacgttctgt	ttcctttaaa	acagagaagc	aagcctctaa	accaaaagaa	780
gtggttttgt	ctatcaaaga	tttgggtggtc	aatgaaaacc	gtggtgttcc	agctgtttaa	840
aatctgtcct	tgagtggtc	tgctggagag	attgttggtg	ttcggggat	tgatggaaat	900
ggtcagtcgt	aactgattca	agccattaca	ggcttctcga	aggttgaatc	tggtagcatt	960
gagctaaaag	gagattcaat	tgtaggcttg	caccacgctc	agattacaga	actaagtgtt	1020
gggcagcttc	cagaagaccg	tcaccgtgat	ggcttgattt	tggaatgat	gatactgaa	1080
aatattgccc	ttcaaaccta	ctataaagaa	ccacatagta	aaaatggaat	tttgtaattat	1140
tcaaatatta	cttcttatgc	taaaaagctg	atggaagagt	ttgatgttcg	cgctgccagt	1200
gaattagttc	ctgcagctgc	actctcagga	ggaaatcaac	aaaaagcaat	tattgtctcgt	1260
gaaattgata	gagatcctga	tctccttata	gttagccagc	caactcgtgg	gttgagtgct	1320
gggtgccattg	agtatatcca	caaacgcttg	attgaagagc	gtgataatgg	caagggtgct	1380
cttgtgtca	gctttgaatt	ggatgagatt	ttaaacgtct	cagaccgtat	tgccgttatc	1440
cacgatggta	agattcaagg	tattgtatca	ccagaaacaa	ccaataaaca	agaacttggt	1500
gtcttgatgg	ctggtggaaa	cttgggaaag	gagaagagtg	atgtctaa		1548
<210>	75					
<211>	939					
<212>	DNA					
<213>	Streptococcus pneumoniae					
<400>	75					
gggaggagaa	caaaaaatgac	agagttggca	aagcaactat	tagagttgac	ctatatgtgtg	60
attggttgct	aatttctcca	tacagcctat	tgtagttata	aagataaaac	aaaccagtt	120
cgacttggga	catctgcatt	ttgactcta	ttgtctatta	cgtttatagg	tggttccat	180
atgccaaata	tgagatttgg	tattattgta	atcctattat	cgctgttaac	attgtttaag	240
caagtcgcta	tcggaacctt	gccatcctta	gatgaaatga	aagccaatat	tgaacttaac	300
agggtgaaaa	ataaaatctt	tattccagtt	atgctgatgg	caatacttgc	gttggtctta	360
gcgcaaatga	ttccagaatt	tagcaagatt	togattagcc	ttgccgcctt	gtttgtctaca	420
atttctgttc	ttgtgattac	caatagtcac	cctaagagtc	tgttatcaga	aaataatcga	480

atgactcagc aagtttcaac aagtggtgatt gtctctcaat tattaggggc tttgggggct 540  
 atttttactg tagcagggtg tggatgatgt atctctcacc tgattagcgg tattgttccct 600  
 tcagatagtc gctttatagg agttttggcc tatgttcttg gaatgggtct attcacaatg 660  
 attatgggaa atgcttttgc agcattcacc gttattacag cagggtgttg agttcccttt 720  
 gtatttgctc tgggagctaa tccaattgtg gctggtgctc ttgccatgac agcagggttat 780  
 tgtgggacct tattgacccc aatggctgct aattttaacg ctctaccacg agcattgatg 840  
 gatatgaaag atcagaatgg cgttataaag gctcaagcag gtgttgctct agtaatgatt 900  
 gttattcaca tattcttaat gtactttctc gcattttag 939

<210> 76

<211> 1113

<212> DNA

<213> Streptococcus pneumoniae

<400> 76

ctcatgtttc gtgaaataa attatttttt tggaccacag aaattttact ctttaaccac 60  
 atcttttacc tatggagaca gatgggtctc ttgattaacc cttttgttag cgtgcttaat 120  
 acaattatga ttccattttt attagggggc tttttttatt atttgacaaa ccctattggt 180  
 actttcttaa ataaagtctg taaactcaat cgtttgcttg gtattttaat taccttgtgt 240  
 actttggtct ggggaatggt catagggtgt gtctatctct tacctatttt gattaatcag 300  
 ttatctagtt tgattatct tagtcaaac atttatagtc gagtacaaga cttaatcata 360  
 gacttatcta attatctgc gctccagaat ttggatgtag aagctacaat tcagcagtta 420  
 aacttatcct atgttgatat tcttcaaaa atcctaata gcgtatcaaa tagtggtggg 480  
 agcgtcttgt cagctcttat cagtactgtt ttgattttga ttatgactcc agtttttttg 540  
 gttttattct tattatagtg acataaatc ttgccatgc ttgaagaac gattctaaag 600  
 agggatcgct tgcattatgc aggtctatta aagaatttaa atgcgacgat tgctcgctat 660  
 attatgggag ttctgatga cgcaatcatt atagggtgtt tggcttatat tggctatagt 720  
 attattggtt taaaatatgc tttagttttt gccatttttt ctggtgtagc caatttaatt 780  
 ccttatgttg ggccaagtat tggtttgatt cctatgatca tcgcaaatat attcactgat 840  
 ccccatagac tgctgattgc agtgatttat atgcttggtg ttcagcagggt agatggcaat 900  
 atcttatatc ctogaatcgt aggaagtgtt atgaaggttc atccaatcac gatttttagt 960



```

ttacttttgt tgtcaagcaa tatctatggt gtagttggaa tgattgtcgc agtgccaacc 1020
tattctatct tgaagaaat ttctaagttc ttatcccatt tgtagaaaa tcataaaata 1080
atgaagaac gagaaagaga attagctaag taa 1113

<210> 77
<211> 1995
<212> DNA
<213> Streptococcus pneumoniae

<400> 77
attagtatgt ttcgattaac caataagtta gcggtatcga acttgattaa aaaccgcaaa 60
ctctactatc cctttgcact ggctgttctc ttggcagtc ccatcaccta tctcttttac 120
tcctaacct tcaatccaaa gattcgggaa atccgtggag gaaccacat tcaagcaaca 180
cttggtattg gtatgtttgt cgttaccett gcgtcagcca ttatcgctct ctatgccaat 240
agttttgtca tgaaaaaccg ttccaaggaa ctgggtatat atggcatggt aggtctggag 300
aagcgccatc taatcagtat gacctttaag gagttagtgg tatttgggat tctaactgtt 360
ggagcgggta tcgggtattg agccttgttt gacaagttaa ttttcgcttt cctgctcaaa 420
ctaatgaaac tgaaggttga gctggttgct accttccaaa tgaatgttgt cattgcagta 480
cttgttgtct ttggattgat ttctctagtc ctcatgttcc tgaatgtctc tgaatcgcc 540
cgtatgaatg cctccagct ctgcgtgag aaagcaagcg gagagaaaag aggtcgcttc 600
ctacctctcc aaacgattct tggttccata agtttaggga ttggctatta tcttgccttc 660
acggtaacgg atcctcttac agccetaaca actttcttcc tagctgtttt gctgggtatc 720
tttggtaact atctattggt taatgcaggg attacagtct tctacaaat cttaagaaa 780
aacaagaaat actattacca acctataaac ctcatatctg ttccaactt gattttccgt 840
atgaagaaaa atgcggttgg actagcaacc atcgctattt tgtcaacaat ggttttggtg 900
accatgtcag cagcgacaag ctttttcaat tccgcagaaa gctttaaaaa agttctaat 960
cctcatgatt ttgggggttc agggcaaaat gttgaaaaa aagatttggg caaactcttg 1020
agccagtttg caagtgcaca aggttatagt gtcaagaga aagaagtact tcgttacagt 1080
aacttttgga ttgcaaatca agaaggaaac aagttaacta tttttgaaa aggacaaaac 1140
cgtgtccaac ccaacaacagt tttcatggta ttgacaaa aagattatga aaatatgact 1200
ggtcaaaaac tgtctctatc aggaatatgag gtcgggtctct ttgcaaaaa tgacggcagt 1260

```

```

aaaggacaga aagctctaac tctaatgat catcaat ttt ctgtcaaaga agaatttaat 1320
aaagatttca ttgtgaacca tgttccaaat aagtttaata tcttgactac tgattacaat 1380
taccttgttg ttctgtgatt acaagccttt ttggatcaat tcccgattc ggctatctat 1440
aatcagtttt acggtggtat gaatgtaaat gtcagtgaag aagaacaact caaggctcgt 1500
gaggagtatg aaaactacct caatcaat ttt aatgctcaat tagacacaga aggtagctat 1560
gtttatggta gcaatctagc agatgctagt tctcagatga gtgccctctt tgggtggtgtc 1620
ttctttatcg gtattttcct atccattatc tttatggtcg gaactgttct ggtcatctac 1680
tacaacaaa tttctgaagg ctacgaagac cgtgaacgct ttattatctt gcgaaaagtc 1740
ggtttggacc aaaagcaaat caagcaaacc atcaacaaac aggttttaac tgttttcttc 1800
cttcttttgc tctttgcctt catacatctc gcctttgcct accatatgct tagcctgatt 1860
ttaaagtgta ttggtgtact ggatacgact atgatgttga ttgtgacctt gttcatctgc 1920
gctatcttcc tcactgccta tgtgctgatt ttcattgatta cttcaagaag ttatcgcaag 1980
attgtgcaaa tgtaa 1995

```

<210> 78

<211> 1290

<212> DNA

<213> Streptococcus pneumoniae

<400> 78

```

ggacatttta gaagaagagg aaggaaaaa atgagtcgtt tactagt tttat tggttgtggg 60
ggcgttgccc aagttgctat ttcaaagatt tgtcaagata gcgaacatt tacagagatt 120
atgattgcta gccgtacca gtcaaaatgc gatgacttga aagcgaagct agaaggcaaa 180
acaagtacta aaattgaaac tgcagcactt gatgctgaca aggttgaaga agtgattgcc 240
ctgattgaaa gctacaacc agaagctgtt ttgaatgtag ctctgcctta tcaagattta 300
accattatgg atgcttggtt ggcaacaggt gttcactata tcgatacagc caactacgaa 360
gcagaagaca cagaagacc tgagtgccgt gctatctacg aaaaacgttg taaggaaactt 420
ggttttacag cctactttga ctactcatgg cagtgggcct atcaagagaa attcaaagaa 480
gcaggcttga ctgctcttct tggttctggt ttgaccocag gtgtaactag tgtcttttca 540
gcttatgccc tcaaacacta ttttgatgaa atccattata tcgacatttt agactgtaat 600
ggcgttgacc acggttatcc atttgcaacc aactttaatc cagaatttaa tctccgtgag 660

```

```

gtttctgcgc caggttctta ctgggaagat gggaaatggg tcgaagtcga agctatgtct 720
atcaagcgtg agtatgattt cctcaagtt ggacaaaaag atatgtatct cctccaccat 780
gaagaatcgc aatcattggc caagaacatt ccagggtgtca aacgcattcg tttctttatg 840
acttttggtc aatcttactt gacgcacatg aaatgtcttg aaaatgttgg actccttctg 900
acggatacca ttaactttaa cggccaagaa attgttccaa ttcaattttt gaaagccttg 960
cttcagatc ctgccagtct tgggccacgt acagtcggaa aaaccaatat tggatgtatc 1020
tttacagggt tcaaaagcgg tgtcaaaaag actatctata tctacaatgt ctgogaccat 1080
caggaatggt acgcagaggt tggttcgcaa gctatttctt atacgacagg agttccagcc 1140
atgattggga caaaattagt catgaacgga acttggaac aagctggagt gtataacctt 1200
gaggagttag atccagatcc attcatggaa gctttgaatg agtatgggtt gccatgggtt 1260
gtggttgaaa atccacaaat ggtggactaa 1290

```

```

<210> 79
<211> 669
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 79
tctaagagag gagaaaatat ggaagcaatt atcgagaaaa tcaaaagata taaaatcatc 60
gtcatctgta ctgggtctggg ctgtcttgta ggaggatttt tcctgctaaa accagctcca 120
caaacacctg tcaaaagagac gaatttcgag gctgaagttg cagctgtttc caaggactca 180
tcgaccgaaa aggaagtga aaggaagaa aaggaagaac cccttgaa caagatctaact 240
acagtagatg tcaaaagtgct tgtcaaatcg ccagggtatt atgacttgcc tgtaggtagt 300
cgagtcgaatg atgtctgttc gaaggctggg ggcttgacag agcaagcaga cagcaagtcg 360
ctcaatctag ctcaagaagt tagtgatgag gctctggttt acgttcttac taaggaggaa 420
gaagcagtta gtcaacagac tgggtcgggg acagcttctt caacaagcaa ggaagaagaag 480
gtcaatctca acaaggccag tctggaagaa ctcaagcagg tcaagggact gggaggaaaa 540
cgagctcagg acattattga ccatcgtgag gcaaatggca agttcaagtc agtagacgag 600
ctcaagaagg tctctggcat tgggtggcaa acaatagaaa agcttaagaa ctatgttaca 660
gtggatttaa 669

```

<210> 80  
 <211> 1524  
 <212> DNA  
 <213> *Streptococcus pneumoniae*  
  
 <400> 80  
 aaaagcttaa agactatggt acagtggatt aagaattttct ctattccctt aatttacctg 60  
 agttttctat tacttttggct ttattacgct attttctcag catcttatct tgccttggty 120  
 ggctttgttt ttctgctagt ctgtctcttt atccaatttc cgtggaaatc tgcctggtaaa 180  
 gttctaataa tttgcggaat ctttggatgt tggtttggtt ttcaaaattg gcaacagagt 240  
 caagcgagtc aaaatctggc ggattctggt gaaagggtag ggattttgcc tgatactatt 300  
 aaggtaaatg gtgatagttc atcctttcgt ggcaagtcta acggtcgtgc ttccaagtc 360  
 tattataaac tccagtcga ggaggagaaa gaagcctttc aagctttaac tgacctgcac 420  
 gagataggac tagaaggaa gctttcggag ccagaagggc agagaaattt tgggtggcttt 480  
 aattaccaag cctatctgaa gactcaggga atttaccaga ctctcaatat caaaacaatc 540  
 cagtcacttc aaaagattgg cagttgggat ataggagaaa acttgctccag tttaactcga 600  
 aaggctgtgg tttggattaa gacgcacttt ccagacccta tgggcaatta catgacagga 660  
 ctcttctggy gacatctgga caccgacttt gagagatga atgagcttta ttccagtcta 720  
 ggaattatcc acctctttgc cctatctggc atgcaggtag gttttttcat gaatggattt 780  
 aagaaacttc tcttgcgatt gggcttgacc caagaaaagt tgaatggct gacttatccc 840  
 ttttccctta tctatcgagg actaactgga ttttcagcat cggttattcg cagtcctctg 900  
 caaaagctac tggctcaaca tgggggtaag ggcttggata attttgctt gacggtgctt 960  
 gtcctcttta ttgtcatgcc aaactttttc ttgacagcag gaggagtctt gtcctcgctt 1020  
 tatgctttta tctcgacct gaccagcaaa gaaggggagg ggtcaaggc tgttactagt 1080  
 gaaagtctag toactctctt gggcatattg ccattctat ccttctatct tgcggaattt 1140  
 caacttgggt ctatctcttt gacctttgtc ttttctttc tttttgactt ggtctcttta 1200  
 ccgctcttgt ctatcttatt tgtcctttcc tttctctatc cagtcattca gctgaacttt 1260  
 atccttgaat ggttagaggy cattattcgc ttggtctgc aggtggcaag gagaccactt 1320  
 gtctttggtc aaccaaacgc atggctttta atcttattgt taatttctt ggctttggtc 1380  
 tatgatttga ggaataacat taaaggatta acagtattga gtttattgat tacaggtctc 1440  
 ttttctcta ccaagatcc actggaaaat gaaatcacca tgctggatgt ggggcaagga 1500

gaaagtatatt tctacgggat gtaa 1524

<210> 81

<211> 261

<212> DNA

<213> Streptococcus pneumoniae

<400> 81

aataataggg attttaagga gtttgatag tataacctat tattaacat tttattagta 60

ttatctgttg tgatttgat tgcaattttc atgcaaccaa ccaaaaacca atccagcaat 120

gtatttgatg ccagttcagg tgatttgatt gaacgcagta aagctcggg ttttgaagct 180

gtaatgcagc gtttgacagg gatttttagtc tttttctggc tagccattgc cttagcattg 240

acggtattat caagtagata a 261

<210> 82

<211> 867

<212> DNA

<213> Streptococcus pneumoniae

<400> 82

aatagaaata gttggaggaa atatatgctc tcatggttag cagcggtat taaagggatt 60

gtaattgctc ttggatttat cctaccggga atttccggag gggttctagc agcaatctta 120

ggaatctatg aacgaatgat tggctttctg gcccatccct ttaaagactt taaagaaat 180

gttttgtact ttattccagt tgccatcggt atgcttctgg gaatcggctt attttctac 240

ccgattgaat acctgcttga aaattatcag gtttttbtat tatggagctt tgcgggagct 300

attatcggtg cagtctctag cctcctcaaa gaatcaactc gagaatctga ccgagacaag 360

attgatttag cttgggtatg gacaaccttt atcatttctg gattaggact ctatgcctta 420

aattttgtcg ttggaacctt aagcgccagc tttcttaact tcgtcctagc aggcgactta 480

ttggcccttg gcgtcttggt tcctggcctc agcccatcaa atttactttt gattttggga 540

ctctatgctc ctatgttgac tggttttaaa acttttgatt tcttgggaac cttctttcog 600

attggaattg gtgcagggtc aactctcatc gttttttcaa aattgataga ttatgcctta 660

aacaactacc actcacggt ctatcatttc atcatcggtg tcgtcctatc aagtaccctt 720

ttgatcttaa ttccaaatgc aggaacagct gaaagtatcc aatacacagg acttttcaatt 780

gtcgggtatg tcatcatcgc cttcttcttt gcgctgggaa tctggcttgg tatttggatg 840

agtcaattgg aggataata taaataa 867

<210> 83  
 <211> 636  
 <212> DNA  
 <213> *Streptococcus pneumoniae*

<400> 83  
 atcatgtttt acttgagttt gtcaaggatt gctttaagct cctctactag ttagttttct 60  
 gtctctgctg agccattttt ttctttcacg aaatcaaggg ttctctggag aaggttttgg 120  
 gctttggcaa ggactttttt atocgctttt totgcatcta gctgtcctag aaccttgatc 180  
 aattccgtgc ttaattgctg gattttctgac tctttcttac ggcgaatcag ccagaaggca 240  
 atcacgccta ggaggccaag tagactgacc acaatcactc ctgccggaac tgagtttggt 300  
 tcagtcactc tatctgaatc ctactatctc tccgttcctt gttttgcac cttcttggtc 360  
 tgtgcaggct tgctgtcgtc agcatttgct ttcacatctt tgagagagtc caaggcagcc 420  
 cagccttcac agactctact gcagtatgca gaccttactc tgcaaggca ctatcttccg 480  
 gagctttttg agcatctagg aggacagcct tgggtgcac gattttcgga tcagatactg 540  
 ttgccaaagc tttcaagcgt tggcttaact cttgactcaa ggcacgaagt tcagacttgt 600  
 caacttgctc ttgagcttgt gtgctcgttg agctag 636

<210> 84  
 <211> 744  
 <212> DNA  
 <213> *Streptococcus pneumoniae*

<400> 84  
 aataggatta gaattattaa gaaagttggt tctttattgg aaacgatagt atttttaatc 60  
 tctgttttct tagcagggtt tttatccttt ttttctcctt gtatttttcc tctctcgcca 120  
 gtctatgctg ggatttttatt ggaatgatcaa gaaagtgcaa aaagcttttc tttgtttggg 180  
 agaaagggtc tctggtcagg cttgattcga acactttgct ttatcgtggt tatctctctc 240  
 attttcttta ttctaggctt tgggtgctgt tactttggtc atattctcta tgcaaatgtg 300  
 tttcgatatg gcatgggagc tattattatc attttgggtc ttcaccagat ggaaattttt 360  
 catttgaaga aattagaagt tcaaaaaagt tttaccttta aaaaatcaga ttctaactgt 420  
 tattggctcag ctttttttact tggattacc tttagctttg gttggacacc ttgtatttgt 480  
 ccagttttta gtctgtgttt agcacttgcg gcttctggag gcaatggcgc ttggcaaggga 540

gcgatttata ctctcattta cactctgggc atggcccttc ctttcttggt attggcacta 600  
 gcttcaggtc tagtcatgcc atatttttagt aaaatcaagc gtcatatgat gctactaaag 660  
 aaaattgggtg gtttctcat tgttttaagt ggaattttgt tactattagg acaagtaaat 720  
 gttctagctg gaatttttga ataa 744

<210> 85  
 <211> 936  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 85  
 atgggttaagg agttgttcat gaagaaacaa aatttatttt tagtctgtt aagtgtcttt 60  
 cttttgtgct tgggggcttg tggtaaaag gaaagtcaga caggaaagg gatgaaatt 120  
 gtgaccagtt tttatctat ctacgctatg gtttaaggag tatctggtga cttgaatgat 180  
 gttcggatga ttcagtcagg tagtggtatt cactcctttg aaccttcggc aatgatatc 240  
 gcagccatct atgatgcaga tgtctttgtt taccattctc atacactega atcttgggca 300  
 ggaagtcttg atccaaatct aaaaaaatcc aaagtgaagg tcttagaggc ttctgaggga 360  
 atgaccttg aacgtgtccc tggactagag gatgtggaag caggggatgg agttgatgaa 420  
 aaaaagctct atgacctca cacatggcta gatectgaaa aagctggaga agaagcccaa 480  
 attatcgtg ataaactttc agaggtggat agtgagcata aagagactta tcaaaaaaat 540  
 gcgcaagcct tttatcaaaa agctcaggaa ttgactaaga aattccaacc aaaatttgaa 600  
 aaagcgactc agaaaaacatt tgtaacacaa catacagcct tttcttatct agcgaagaga 660  
 tttgggctta atcaacttgg tattgcagggt atctctcctg aacaagaacc aagtcacaga 720  
 caactaacag aaattcagga atttgttaag acctataagg ttaaagcat ttttcagaa 780  
 agtaacgctt cttcaaaagt agctgaaact cttgtcaaat caacaggtgt gggctcttaa 840  
 actctgaatc ctttagagtc agaccacaa aatgacaaga cctatttga aaatcttgaa 900  
 gaaaatatga gtattctagc agaagaatta aagtga 936

<210> 86  
 <211> 390  
 <212> DNA  
 <213> Streptococcus pneumoniae  
 <400> 86

aaggaaaaca gtatgttaaa aaatctaaaa tcgttcttgc ttcgaggaaa tgttattgac 60  
cttgcgtcgc gtgttgtaat tgcctctgct tttggtgcta tcgttacttc acttgtaaac 120  
gacattatca ctctctttat tttaaatccc gctttgaaag ctgctaaagt tgaacgtatc 180  
gctcaacttt cttggcatgg agtcggctat ggtaacttct taagtgcctat tatcaatttt 240  
atctttgtgg gtaccgccct ctctctttat atcaagggca ttgaaaaagc acagaagctg 300  
actggcataa aggaagaaaa aactgacgaa aaaaaaccaa ccgaattgga agtccttcaa 360  
gaaataaaag ctctccttga gaaaaataa 390

<210> 87  
<211> 1023  
<212> DNA  
<213> Streptococcus pneumoniae

<400> 87  
gaaatgcattg caaaaatgcg aaataaaaaa caaataaacc tag\_ataat ttttataatc 60  
tgcctaggtc ttcttattac aatatttttg tcattaaagc ttggaacaaa agaaattaat 120  
atcagagatt ttttagcagc ttttggaatg ggtaatacaa atgatgattt tattaataca 180  
attatatata aaagaatacc tagaactatt tttgcaattt tagcagggtc tagtctcc 240  
ataagcggtg tattgatgca atcagttact agaaacccaa tagctgattc aggtatactc 300  
ggtataaaca caggagcaag tcttagtgta gtaattggtc ttcttttttt aggaatttca 360  
tcaagcataa gccatataag ttttgcaatc attggtggct tagtaagtgc aatttttgta 420  
tacgcgattg ctgtaagcgg aaaagcaggc cttacccta taaaacttgc cttatcagga 480  
acttggtgta gtatggcttt aagcagtttt gtaagttttt taattttacc gaataataac 540  
gtcttagaca aatttagatt ttggcaataa ggtagccttg gagcagctac attatctct 600  
atatctacac tactaccttt tataatttta ggtcacttga tagctatatt tatttcata 660  
gatttaaacg ctttagctat ggggatgaa atggctgttg gtcttgaggat taatgttaat 720  
aggataagat cacttgcaat aattgcaagt gtgcttttat gttcaagtat tactgcaatt 780  
ggtggaccta ttggcttcgt aggtcttata gttcctcact tttgtggctt atttataagc 840  
aaagatatag gccaatgac catttcttca tcttttatag gtgcagagct cttgcttata 900  
tgtgatataa tcggccgtat gttaggtaaa ccagggtgaa ttgaagtagg gataattact 960  
gcaataatcg ggggtccagt acttatttat gtaactatga aaatagagg ggtaataaac 1020



taa 1023

<210> 88  
 <211> 1011  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 88  
 ctaatgcaaa atttaattat aggtattoaa aaaagaaaa atagaataac actatatttc 60  
 tcactatttt tattaataat aatcagtcata tcatttttca tttactttat cggagatgaa 120  
 agttattcct tttcaacttt gattaaagtc ttaaatagtg aaactgttcc tggagctagt 180  
 ttttcgatta tggaaattag attacaaaa ttattagcag gaattatagc tggctgggtct 240  
 tttggattgg caggatttat ctttcaact atgttaagaa atcctcttgc aagtcctgat 300  
 ataatcgggtg tcacaagttc ttcctctatt gcacgggtct ttgcatatt ggtattaaaa 360  
 acaaatagtt taactactgg aattatttca ataactgtg gactaacatc atctttaata 420  
 ttatttttac tagctaaaa agatggtttt tcagcagcaa gactgataat attaggtatt 480  
 ggttttcaag ctgtcacaag agcaggcacc tcatttttat tgttgaaagt agcaagatat 540  
 gaattacaag aagttatgag atggctcagt ggctctttat cttttacaaa gttatagatgac 600  
 atacctcttg ttctaatagt aagtattatt gctactatat tagttttatt ttttaataaa 660  
 agactagaaa ttattgaact tgggtgaagaa atagcaatcg gacttggagc aaatcccgag 720  
 ctttcaaggc ttgttttaaat tttttgcgct gtatctttaa ctgcttttcc tacttcaatt 780  
 acaggaccaa tagcttgat atctttttta gctgggtcca tagccttaaa tattggcaag 840  
 aaaagaagcc caatattagc tggattgggt ggaattttac tagttttgtt atcagacata 900  
 ttctctcaaa atatttttacc agctagatat ccagtaggtg ttgtaactgg cttgttaggt 960  
 tcaccatact taatatactt actaataaaa atgaacagga ggaatatata a 1011

<210> 89  
 <211> 936  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 89  
 aataaaggta ggggtgttat gaattgttcc ttgaaatga ataatgtaag tgttcgttat 60  
 gatgacgtaa tagctttaaa agatataact ttacaaataa ataagggaga tttcattggc 120  
 ttattagggt caaatgggtgc aggtaaatct acgttaatta attctattgt aggttttcaa 180

gagatttatt taggagaaat agagtattgt gataaagatt tgatagttag ttctcaacct 240  
 ttgtctcatt taggctttac tcctcaaacc acagtaattg atttttatac tactgtgaag 300  
 gacaatgtaa tattggggct gaaccttgct ggaaagtttg ggaataatgc tgagaagttg 360  
 tgtcaaatag ccttagaaat tgttgggtta gctgataaaa aaaataattt ggtagaaca 420  
 ttgtcagggt gacaactgca acgcgtccag attgctagag caatagctca taatccagat 480  
 ttttatattt tagatgaacc taccgttggt ttgatactg aatctgcga aaaattttta 540  
 atgtatttaa aagataagag ttggaagga aaactatta tcatacttc acatgacata 600  
 aatctactcg aaaagttttg taaaaaata ctttttttac aaaatggctc catatcattt 660  
 ttgtgtgata tgcgtgactt ttagataat tcaactatca aattaaattt ttcaatgcag 720  
 aatagaattt ctatgatca aattgaattt ttagaaaatt ttagatttaa agttcacatc 780  
 gaagataatg atagttttac aatagaagtc cctatagaag aaaagatctt agatgttatc 840  
 aatgaggtag gaaaagcatg tgaatttaa aacttttcaa caagtaaatt aaccttaca 900  
 gaaagttatt tgcaagaat aggaggagaa aatga 936

<210> 90

<211> 846

<212> DNA

<213> *Streptococcus pneumoniae*

<400> 90

attaacctta caagaaagtt atttgcaaag aataggagga gaaaatgaa ggctgatcaa 60  
 ttaaggcaca aatcggaactt aggtttaaga ggtctagcga ttattgctaa aaatgagatt 120  
 attgcttttt ttagaagtaa aggtttaatt atttctcagt ttctacaacc aatcttatat 180  
 gttgttttta taataatagg attaaattct tcgataaga acattcagtt taatgatata 240  
 aaaacctctt atgcagaata tacaatcatt ggtgttatag ctttattgat aatggggcag 300  
 atgactcaag ttattttatag ggtgacaata gataaaaaat atgggctact tgctcttaag 360  
 ttatgcagtg gagttcgtcc ttatatattt attttaggga tgagtatcta ttctatatta 420  
 ggggtgtag ttcaagaat tatttatat ataattacgt tagcgtttga gataaatatc 480  
 gcaatggata gattttttta tacagttttg ttatctattg ttgttttatt attttgggac 540  
 tccttgcga ttttacttac aatgtttatc aatgattaca gaagacgtga tattgtaata 600  
 cgtttgttac taacaccgct tggttttaca gtcctgtgtt tctacttaat agattctgct 660

cctagtattg tgagatggat tggtcagtta aatcccttaa cttatcaatt aactattttg 720  
 agaaactttt attttaaaaa ttcaacaact ttggaattag tttctttatt gttacatca 780  
 ttacttgtcc ttatatotgt atcttttatt ataccaaaga taaaattgat actgatagaa 840  
 agataa 846

<210> 91  
 <211> 1038  
 <212> DNA  
 <213> *Streptococcus pneumoniae*

<400> 91  
 aatatgttaa aggaataaaa aaggagaaac agaataaaa ataaacgttt aattggaatt 60  
 attgctgcat tagcagtcct agtagcagga agcttgattt attcttcaat gaataaatca 120  
 gaagctcaga ataataagga tgagaagaaa ataaccaaga ttgggtgtgct tcaattttgtg 180  
 agccatccat ccccttgattt gatttataaa gggatccaag atggacttgc agaagaagga 240  
 tataaagatg atcaagttaa aattgatttt atgaactcag aaggtagcca aagtaagggtt 300  
 gcgacaatga gtaacaattt ggttgcaaat gggaatgacc ttgtgggttg tatcgcaaca 360  
 ccagcagccc aagggttggc tagtgcaaca aaagacctac cgtttatcat ggccgctatt 420  
 acagacccaa ttggtgctaa cttggttaaa gatttgaaaa aaccagggtg caacgtttaca 480  
 ggggtatctg accacaatcc agctcaacaa caagttgaac tcatcaaggc tctgacaccg 540  
 aatgtgaaaa caatcgagc tctttactca agtagcgaag acaattcaa aacacaggtc 600  
 gaagaattta aggcctatgc tgaaaaagca ggtctgacag tggaacatt tgcagttcct 660  
 tcaacaaatg aaattgcctc aactgtcact gttatgacta gcaaggtaga tgctattttg 720  
 gttccaattg ataaccatc tgcacagga tttccaacgg ttgtctctag caatcaaaat 780  
 tctaagaac caatttatcc cagtgcgaca gctatggtag aagttaggtg tttggcatca 840  
 gttgtaatg accaactatg ccttggtgtg gcaacaggtg aaatgattgt gcaagtcctg 900  
 aaagggtcaa aaccagccga taacccagtc aatgtctttt caactggtaa gtcagtcac 960  
 aataaaaaaa tagcacaaga actagggtatt actattcctg agtctgttct caaagaagca 1020  
 ggacaagtca togaataa 1038

<210> 92  
 <211> 792

&lt;212&gt; DNA

<213> *Streptococcus pneumoniae*

&lt;400&gt; 92

```

gcaaccaatc ttgaaaggag ccaagttaag caaatgacag caattgtaga attaaaaaat      60
gcaacccaaa tcgttaaaaa tggctttgat gaagaaaaga ttattttaaa tgatgttttc      120
ttagaaattt ttgaacggga ctttatcacg attttggcg gaaatgggtgc tggaaaaatca      180
actctcttta acactatagc agggacctta tcactaacta gtggaactat cegtatttta      240
ggatgaagatc tcactaagtt ttcacccgag aagcgtgccca agtacctgtc tcgtgtcttc      300
caagatccaa agatggggac agtccccgt atgacgggtcg ctgaaaatct ttaaatcgcc      360
aagtttcgtg gtgaaaagcg tggattgtta ccacgaagct tgactagcta taaggatgaa      420
tttcaggcaa ccatgaaaa agtaggaaat ggtcttgaga aacacttgaa tacaccgatt      480
gagttcttat caggtggaca aagacaggct ttgagtctct tgatggcaac cttgaagcga      540
cctgaattac tctgtttaga tgagcatact gctgccctgg atccaaagac tagtgtgtgt      600
ttgatggaat tgacagatga atttgttaag aaagatcagc taacagccct tatgattact      660
catcatatgg aagatgctct caaatcgggc aatcgcttga ttgtcatgaa agaaggacga      720
attatccaag atttaaacca agaagaaaaa gcaaaaatga aaatctctga ttattatcaa      780
ctctttgaat aa                                         792

```

&lt;210&gt; 93

&lt;211&gt; 741

&lt;212&gt; DNA

<213> *Streptococcus pneumoniae*

&lt;400&gt; 93

```

aaagaaaatg gtacaatatt tctaagagaa aatacaatgg gaggtaaaat gaggttatta      60
cctataagaa aaatatcacg tcagtcataa aggttagcac tttttttgac gttttgtgct      120
ggatatgtgg atgcttacac ttttattgtt cgcgggaata ccctttagtc tggacaaact      180
ggaaatgttg tctttcttct agtagaatta attaaaaata atgtttcgga tgttagggac      240
aagggttctca cttgtctagc gtttatgatg ggagtccttt tattaacgat ttataaggaa      300
aaattgagaa ttgtgaaaaa acctattctg tccttgatcc ccttggcaat cttatcaatc      360
attattgctt ttgtgccgca aactgtggat aatatctatc tagtgccgcc cttggccttc      420
tgtatgggac tggtgacaac tgcttttggg gaagtgtcgg gtattgccta taataacgct      480

```

ttatatgacag ggaatatcaa acggaccatg ctggcttttg gagattattt cgaaccaag 540  
 cacactccctt ttttgcgtga aggattcata tttgttagcc tgettagtag tttgtccctt 600  
 ggcgttgtct tttcagccta tttgacgatt ttctatcatg aaaagaccat tcttggtggt 660  
 cctattatga tgagcgtttt ttacctcagc atgctttttg cctcttgga gaaaaaagta 720  
 aaagaaaaag ctctatttta g 741

<210> 94  
 <211> 864  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 94  
 gaaaagagggt gtcctatgat taaaaaatt tacccattt ttaccattt actagggtct 60  
 gctatttatg cttttggact gacttattt gtatttccc atcatctctt tgaaggagggt 120  
 gcgacaggca ttacctcat caccttttat ctttttaaaa tcctgtttc cctcatgaac 180  
 ctgctgatta atattccctt ttctaccta gcttgaaga tttttggagc caaatccctc 240  
 tattctagtt tactaggaac cttagctttg tccggctggt tagctttttt tgagcatatt 300  
 ccccttcata ttgatctta aggtgattta ctaatcacag ccttatagc gggaatccta 360  
 ttgggaattg gccttgaat tatttttaat gctggaggta caactggcgg aactgatatt 420  
 ctagctcgta ttctcaacaa atacactcat atatccatag gaaaactgct ctttatctta 480  
 gatttttgta ttctcatggt gattctccta atcttcaagg atttgagatt ggtttcttac 540  
 acgcttttgt ttgattttat tgtttctgt gttattgatt tgattggtga aggaggatat 600  
 gccggcaaa gctttatgat tatcacaaaa cgtcctgacc aacttgctaa ggcgattaat 660  
 gatgacctcg gaagagggtg tacttttatt tctggccaag gctactatag taaagaaaat 720  
 ttgaaatca tetactgtat tgcggaaga aatgaaattg tgaaaacgaa ggaatgatt 780  
 catgaaatg atcctcaagc ctttataact attacagaag ccatgaaat cctagagaa 840  
 ggcttcacct ttgaaaaaga ataa 864

<210> 95  
 <211> 300  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 95  
 aaattatttg gaggaatcat taacatggca aacaaacaag atttgatcgc taaagtagca 60

```

gaagctacag aattgactaa gaaagactca gcagcagcag ttgaagctgt atttcgacga 120
gtagctgact atcttcgcgc tggtgaaaaa gtccaattga tcggttttgg taactttgaa 180
gttcgtgagc gcgcagaaacg taaaggctgc aaccacaaa ctggttaaaga aatgacaatt 240
gcagcttcta aagtaccagc attcaaagct ggtaagctc ttaaagacgc tgttaaataa 300

<210> 96
<211> 1095
<212> DNA
<213> Streptococcus pneumoniae

<400> 96
tatacttact tatggagaaa atacatgaaa cgtgagattt tactggaacg aatcgacaaa 60
ctaaaacaac tcattgccctg gtatgttctg gaatactacc aatctaagct ggctgtgccc 120
tacagtttta caacctgtga cgaatacctt aaggaatatg accgattttt cagctgggtt 180
ttggagctcg gtatttcata cgctgataaa atatccgata ttcttttata agtttttgaa 240
aatatgtcta agaaagacat ggaatccttt atcctttata tacgtgaacg tcccttgctg 300
aatgctaata caacaaaaca aggtgtttca cagacaacta tcaatcgaac cttatcagca 360
ctttctagtc tttaacaagta tctaaccgag gaggttgaaa acgatacggg ggaaccttat 420
ttctatcgta atgtaatgaa aaaagtcttc accaagaaaa agaaagaaac ccttgctgcc 480
agagctgaaa atatcaagca aaaactcttt ctaggatgat aaacagaagg ttttctaact 540
tatatcgatc aagagcacc cacaacagctt tcaaatcgag ctctctcatt atcaacaaa 600
aataaagaac gagatttagc cattattgcc ctctctcttg catctgggtg tcgcttatct 660
gaagctgta atctagatct aagagatctc aatctaaaa tgatggttat tgatgttact 720
cgaaaaggtt gcaaacgtga ctacgtcaat gtcgctgctt ttgctaaacc ttatttagag 780
aattatcttg ccattcggaa tcaacgctat aaaacggaaa aaacagatac agcccttttt 840
ttaactctct acagagggtg tcttaactgt atcgatgctt ctacggttga gaaaatgggt 900
gctaataact cagaggattt taaagtcgt gtaacacccc ataaactcgc ccatacacta 960
gcaactaggc tctatgatgc gactaaatca caagttttag tcagtcacca actaggacat 1020
gctagcacac aagtcactga cctctatacc catattgtta gtgatgaaca aaagaatgct 1080
ctggatagtt tatga 1095

```

<210> 97  
 <211> 405  
 <212> DNA  
 <213> *Streptococcus pneumoniae*

<400> 97  
 ctgagggctg caccgcagc gatcccccata ccaccacctt cgataccatt ggcaccaagg 60  
 ttcccagcat caaggctcagc gatatgcata gatccacctt tccctttaca ggttccagtg 120  
 tattttaccaa ggatttcagc catcattccg ttgaggtcaa tccctttagc aatagcttgc 180  
 ccggtgccac ggtggtttga ggtaatcaga tcattctggat tgagagctaa catagcccc 240  
 acgtagctg cctcttcacc aacagaaaag tgcgtcattc ctggcacttt cctcttcttt 300  
 actaattgtg caatttttaa gtccatgcga cggatttctt ccatcttacg gaacatttct 360  
 agcaaaagat tttttatcaa agttgacatc ttcttgcctt tctaa 405

<210> 98  
 <211> 1716  
 <212> DNA  
 <213> *Streptococcus pneumoniae*

<400> 98  
 actctacaag agaggagttc aataatgaat aaactaattg catttatoga gaaaggaaag 60  
 cctttctttg aaaaactatc tcgtaatatc tatcttcgtg ctattcgtga tgggttccatt 120  
 gcaggtatgc ctgttattct cttctcaagt atctttatct tgattgcctt tgtaccaaac 180  
 tcatggggct ttaaatggtc tgatgaagtt gtacctttc tgatgaaacc ttagctctat 240  
 tctatgggta ttctggctct cttggtagct ggtacaacag ctaagtcatt gactgactca 300  
 gtaaacccga gcatggaaaa aaccaatcaa atcaagtata tgtcaacatt gttggcgca 360  
 attgttggtt tgttgatggt ggcagctgat cctatcgaaa gtggtctagc tactggattc 420  
 ttggggacaa aaggtttgct ttcagccttc cttgctgcct ttgttactgt agccatctat 480  
 aaggtttgtg ttaagaacaa cgtcactatt cgtatgctg acgaagtcc accaaatctc 540  
 tcacaagtct ttaagatgt gattccattc actctatctg ttgtttctct ttagctctct 600  
 gacttattag cacgttattt tgttggttct agtgtggcag aatcaatcgg taaattcttc 660  
 gcaccactct tctcagcagc agacggatcc cttggtatta ccattatctt tgggtgcctt 720  
 gcctctctct ggtttgttgg gattcatggt ccatctatcg ttgaaccagc tatcgagct 780  
 attacctatg ccaatgccga agttaacttg aaccttctcc aacaagggat gcatgcagac 840

aagattctta cttctggtac acaaatgttt atcgttacca tgggtggtac aggtgcgaca 900  
 ttggtcgttc catttatgtt catgtggttg acaaaatcga aacgtaaccg tgcaatcgga 960  
 cgtgcttcag tagttcctac cttcttcggt gtaaatgaac caatcttggt ttggtgcacct 1020  
 cttgttttga atccaatctt cttcattcca tttatctttg ctccaattgc aaacgtatgg 1080  
 attttcaaat tctttattga aactcttgga atgaactcat tcactgctaa tctaccatgg 1140  
 acaactccag ctccactagg tctagtcttt ggaactaact tccaagtgtc atcatcatt 1200  
 cttgctgccc ttctaactgt ggttgacgtt gtcatttact atccattcct taaggcttat 1260  
 gatgaacaaa ttcttgaaga agaacgttca ggtaagtcta atgatgaatt gaaagaaaaa 1320  
 gttgtgcaa acttcaacac tgcaaaaagc gatgctattc ttgaaaaagc ggggtgctgat 1380  
 gcagcacaaa ataccatcac tgaagaacaa aatgtcctcg ttctctgtgc aggtggagga 1440  
 acaagtggtc tctttgcaaa tgetttgaat aaggcagcag cagaatacaa tgctcctgtg 1500  
 aaagcagcag caggcggtta ttgtgtctac cgtgaaatgt taccagagtt tgatcttgtt 1560  
 atccttggcc ctcaagttgc ttcaaaacttt gaagatatga aagcagaaac agataagctc 1620  
 ggtattaaac tagcgaaaac agaaggcgct caatacatca aattaactcg tgatggaaaa 1680  
 ggtgctcttg cattcgtaca agcgcaattc gattaa 1716

&lt;210&gt; 99

&lt;211&gt; 807

&lt;212&gt; DNA

<213> *Streptococcus pneumoniae*

&lt;400&gt; 99

gagtttatta tgggtttcttc ggaatttatc tcaagatttg aatttgcttg caataagaaa 60  
 gaaagtcttt atagtcaaag caaatattaag tatcgattc gttcgatgtt cgcaggtgca 120  
 tttttaacct tcagtactgc tgcaggtgca gttggggctg acttgattaa taaaattgca 180  
 ccaggtagtg gacgtctcct ctttcattc gtttttgctt ggggcttggc ctacattggt 240  
 tttttgaatg ccgagttggt cacttcaaac atgatgttct tgactgctgg tagtttctta 300  
 aaaaaaatct cttggagaaa aacagctgag attttactat actgtacctt gtccaacctt 360  
 atcggagcct tgatagcagg gtggggcttt gctcattcgg cagcctatgc gaactcgaca 420  
 cagcatagtt tcactctcagg tgtgtttgag atgaagttag gccgctcaaa tgaattggtc 480  
 ttgcttgagg cgattttggc aaatatTTTT gtaaatattg cgattctgtc atttattttg 540



gtcaaagatg gtgggtgccaa actttggcctt gtgttgtcag ctatttacat gtttgtattc	600
ttaacaaacg agcacatttcg ggcgaacttt gcttctttcg cgatttgtgaa attcagtggt	660
gctgcggatt caattgccaa cttcgggtgtt ggaaatatgc ttgcaccatg ggggtgtgact	720
ttcatcgga actttatcgg agggagcctc ttgatgggtc ttccatatgc cttcctcaat	780
aaaaacgaag atacttatgt agattaa	807

&lt;210&gt; 100

&lt;211&gt; 1356

&lt;212&gt; DNA

&lt;213&gt; Streptococcus pneumoniae

&lt;400&gt; 100

gaaataatgc ttgatttact gaaacaaacc atttttaccag gagattttat ctttatcctg	60
atthttgttag gtttcatcct tgtttgtacc ctcttattac tggaataag acgtgataat	120
attcagttga agcaagtcaa tcaaaagggt aaagatttga ttgcaggaga ttattccaag	180
gttcttgata tgcaagggtg gtctgaaac accaatatta ccaataattt gaatgacttg	240
tcggagggtta ttctgtctac tcaggaaaat ctagaacaag agagtaagag gctaaatagt	300
attctgtttt atatgacaga tgggggttctt gcgactaacc gtccgggttca gattatcatg	360
attaacgata cagccaagaa gcaactgggg ttgggttaagg aagatgttct gaatagaagc	420
atthttggaat tgctcaagat agaagaaaac tatgaattgc gtgatttgat tacccaaagt	480
ccagaattgt tgctagattc ccaagatata aatggcgaat atttgaacct tcgagttcgc	540
tttgccctga tacgtcgaga gtctgggtttt atttcagggt tgggtggctgt tttgcattgat	600
acgacggagc agggagaagg aagaacgcga cgaagactct ttgtttccaa tggtagccat	660
gagttacgga ctctcttgac tagcgtaaaa tcttatcttg aagccttgga tgagggggct	720
ttgtgtgaaa ctgttagcacc agactttatc aaggtttctc ttgatgagac caacgtagt	780
atgcgcattg tgacggtatc cctccatctt tcacgtattg ataattgtac cagtcaacct	840
gatgtggaac tgattaaact cactgctttt attaccttta tccctcaatg ttttgacaag	900
atgaagggac aggaaaagg aaaaaatat gagtttgtga gagattatcc catcaattct	960
atctggatgg aaattgatcc agataagatg acgcagggtg tcgacaatat ttaataaat	1020
gctattaagt attcgcaga tgggggttaa atcactgtca gaatgaagac aactgaagac	1080
cagatgattt tatccatttc tgaccacggt ttggggattc ctaagcagga tttaccacgt	1140

atctttgacc gtttctatcg tgtggatcgt gctagaagtc gtgcacaagg tggtagcagg 1200  
 ctaggactgt ctatcgctaa agaaattatc aaacaacata agggctttat ttgggccaag 1260  
 agtgaatacg gcaagggttc aacctttacc attgtactcc cttatgataa ggatgcagtg 1320  
 aaagaagaag tatgggagga tgaagtagaa gactag 1356

&lt;210&gt; 101

&lt;211&gt; 594

&lt;212&gt; DNA

&lt;213&gt; Streptococcus pneumoniae

&lt;400&gt; 101

attcttgcct tattgtctgg tttattgtac catactagtg tatatgcagt taaaaaggag 60  
 attcttgtga atacacggaa aaagacacaa tttatgacaa tgacagccct tttacaggct 120  
 attgcgattt tgattccaat tgttatgcct ttcaagattg tcattccacc tgcttccctat 180  
 actttgggga gccacatcgc tatttttata gccatgttct tgcgcacct gatggcagtt 240  
 ttgtgcatcc tagcctctag ttttggattt ttgatggctg gctatcccat ggttatcggt 300  
 ttccgggctt tttcccatat atcttttggg gctttaggag ctctttacct acaaaaaattc 360  
 ccgataccc tagataaacc aaaatcttcc tggattttca actttgtttt ggctgttgtt 420  
 catgcccttg ctgaagtatt ggctgtgtgc gttttttacg caacttctgg taccaatgta 480  
 gaaaaatatgt tttatgttct attgtacta gttggatttg gtacaattat coactagtatg 540  
 gtagactata cattagcact agctgtctat aaagtgcctc gaaaacgccg ttaa 594

&lt;210&gt; 102

&lt;211&gt; 867

&lt;212&gt; DNA

&lt;213&gt; Streptococcus pneumoniae

&lt;400&gt; 102

attgctatta aaaagtgcct taataatagt atatatagaa gaaaagagga cggggatatg 60  
 aagaacaaaa gaatatitaa agacttccaa gcttcaaaaa tgagttttaa catttacaca 120  
 agcccttgt tagcctttgt tttgtcttc ataggagagt ttgtggcttt tactttgtat 180  
 ggtattggct tgttagctct catcggaact gctagaaatt ttggagaggc tgggtcaaat 240  
 cttgcaagct acttgacagc cttgcatcag agcttgacgg ataaaaacag tgactttcgt 300  
 ttaatttttag gattactggc ctttggtttt attcttaaca ctgtgttcag atggacaaga 360  
 aaagttgaga aaagacctat tcgaaccttg ggattttata gagagaattt cctcagcaat 420

```

cttctgaaag gatttagtct aggcctggca ctttttcttc tgacctgtt aggttttagtg 480
gtcttaggtc aatatcgttt ggaatccatt cacttgaatc cttattctct tgcccttgtc 540
gtctttacta tcccatcttg gattttacag gggacagcag aagaagtggt gggccgtgct 600
tggctacttc ctcaattggc ctcaagaacc aatctaaac tagctattct tatatctagc 660
ctgtttctta cctgcttca tatgggcaat tctggctca cccctctatc tctagtaaat 720
ctctttttat tcggagttgc catggctctt taccttctca aaactgatac agtttggggt 780
gttgacagta ttcatgggtc ttggaatttt gctcagggtc atctctttgg gattttagtt 840
agtggtaaac cgtcagaacg tctctga 867

```

<210> 103

<211> 2193

<212> DNA

<213> *Streptococcus pneumoniae*

<400> 103

```

gagaatatc ggaaaggag actaaaaatg aagaaaaat ttctagcatt ttgtctaatt 60
ttattcccaa ttttctcatt aggtattgcc aaagcagaaa cgattaagat tgtttctgat 120
accgctatg cactttttga gtttaagat tcagatcaaa cttataaagg aattgatggt 180
gacattatta acaaagtcgc tgagattaaa ggctggaaca ttcagatgct ctatcctgga 240
tttgacgag cagtcfaatgc ggttcaagct gggcaagccg acgctatcat ggcagggatg 300
acaaagacta aagaacgtga aaaagtcttc accatgtctg atacttacta tgatacaaaa 360
gttgtcattg ctactacaaa gtcacacaaa attagcaagt acgaccaatt aactggcaaa 420
accgttggtg taaaaaacgg aactgccgct caacgtttcc ttgaacaat caaagataaa 480
tacggcttta ctattaaaac atttgacact ggtgatttaa tgaacaacag cttgagtgtc 540
ggtgccatcg atgccatgat ggatgacaaa cctgttatcg aatatgccat taaccaaggt 600
caagacctcc atattgaaat ggatgggtgaa gctgtaggaa gttttgcttt cgggtgtgaa 660
aaaggaagta aatacagaca cctggttact gaatttaacc aagccttgct tgaaatgaaa 720
aaagtggtta gtcttgataa aattatcaag aaatggactg cttcatcatc ttcagcagtg 780
ccaaactaaa ctactctcgc aggattaaaa gctattcctg ttaaggctaa atatatcatt 840
gccagcgatt cttcttttgc cccttttgtt ttcacaaatt caagcaacca atactggt 900
attgatatgg aattgattaa ggcaatcgct aaagaccaag gttttgaaat tgaatcacc 960

```

```

aaccctgggtt ttgatgctgc tatcagtgtc gtccaagctg gtcaagccga tgggtatcatc 1020
gtcgggtatgt ctgtcacaga tgctcgtaag gcaacttttg acttctcaga atcatactac 1080
actgctaata ccattcttgg tgtcaaagaa tcaagcaata ttgcttctta tgaagatcta 1140
aaagaaaaga cagtcgggtg taaaaacgga actgcttctc aaaccttctc aacagaaaat 1200
caagcaaat acggctacaa aatcaaaacc tttgctgatg gttcttcaat gtatgacagt 1260
ttaaacactg gtgccattga tgccgttatg gatgatgaac ctgttctcaa atattctatc 1320
agccaaggtc aaaaattgaa aactccaatc tctggaactc caatcgggtga aacagccttt 1380
gccgttaaaa aaggagcaaa tccagaactg attgaaatgt tcaacaacgg acttgcaaac 1440
cttaagcaa acgggtgaatt ccaaaagatt cttgacaaat acctagctag cgaattctca 1500
actgcttcaa caagtactgt tgacgaaca acgctctggg gcttgcttca aaacaactac 1560
aaacaactcc ttacgggtct tgggtatcact cttgctctag ctcttatctc atttgctatt 1620
gccattgtca tcggaattat cttcgggtatg tttagcgtta gccatacaa atctctctgc 1680
gtcatctctg agattttcgt tgacgttatt cgtgggtatc cattgatgat tcttgagacc 1740
ttcatcttct ggggaattcc aaacttcac gagtctatca caggccaaca aagccaatt 1800
aacgactttg tagctggaac cattgcctc tcaactcaat cggtctgtta tatcgctgaa 1860
atcgttcgtg gtggtattca ggccgttcca gttggccaaa tgggaagccag ccgaagcttg 1920
ggtatctctt atggaaaaac catgcgtaag attatcttgc cacaagcaac taaattgatg 1980
ttgccaaact ttgtcaacca attcgttata gctcttaaag atacaactat cgtatctgct 2040
atcggtttgg ttgaactctt ccaactcgtt aagattatca ttgctcgtaa ctaccaaggt 2100
ttcaagatgt atgcaatcct tgctatcttc tatcttgtaa ttatcacact tttgactaga 2160
ctagcgaaac gcttagaaaa gaggattcgt taa 2193

```

```

<210> 104
<211> 774
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 104
actagcgaaa cgcttagaaa agaggattcg ttaatggcaa aattaaaaat tgatgtaaat 60
gatttacaca agcactatgg aaaaaatgaa gtcctaaaag gaattacgac taagttctat 120
gaaggagatg ttgtttgtat catcgggtcct tcaggttctg gtaagtaaac ttctctcgt 180

```

```

agcctcaatc ttttagaaga agtcactagc ggtcacatca ctgtgaacgg ctatgattta 240
actgaaaaaa caaccaatgt tgaccacgtc cgtgaaaata tcggcatgggt attccaacac 300
ttcaacctct tccctcatat gtctgtattg gacaacatca cctttgctcc tattgagcac 360
aagttgatga ctaaggaaga agctgaggaa ttgggaatgg agttgcttga aaaggttgga 420
ctagcagata aagctaattc caatccagat agcctatcag gtggtcaaaa acaacgtgtg 480
gccatcgctc ttggcctagc aatgaatcca gacatcatgc tcttcgatga accaacttct 540
gcccttgacc ctgagatggg ttggagacgt cttaacgtta tgaagggaatt ggctgagcaa 600
ggcatgacca tgattatcgt aaacctgag atgggatttg ctgctcaggt tgccaaccgc 660
gttatcttta ctgcagatgg cgagttcctt gaagacggaa cacctgacca aatctttgat 720
aaccacaac accctcgtct gaaagagttc ttagataagg tcttaaacgt ctaa 774

```

<210> 105

<211> 372

<212> DNA

<213> *Streptococcus pneumoniae*

<400> 105

```

ctaggagaag ttatgcgtat tatctatcta attattgggt ttttatcgct gaccttggct 60
attgttgggg ttgttttacc cttgttgctt acaacacctt tccttttggt gtctattgct 120
tgtttctcca gaagttccaa gcgattcgaa gattggcttt atcataccaa gctctatcaa 180
gcatatgtag ctgattttcg tgagaccaag tctattgcgc gtgaacgaaa gaaaaaaatc 240
atcgtctcta tctacgtctt gatgggaatt tctatttatt ttgcacctct tttaccagtc 300
aaaatcggtc tgggtgcttt gacctctttt attacttatt atctcttcaa ggtcattcca 360
gacaaagaat ag 372

```

<210> 106

<211> 555

<212> DNA

<213> *Streptococcus pneumoniae*

<400> 106

```

tactgtacgt tttatcatag aaatttttac tttattttct catcaaatga gatttgcatc 60
aatctcttgt cttacttgcg tttctttctt gctttcttca ttttgtagc catacgcttc 120
atggactggt tcatggcaaa ttcaccaatt ttacctttca aaccgccacc aaacatctgg 180

```

ctcatatctg gcattcctgc tctccgaga gctgataagt caggcatacc gccttgtccc	240
atcattcctt caagggcaga catatccatt cctcccatat ttggcatatt tttagggaagg	300
ttatttggat taatcccat ttgtctcctc attttattca tatcccaga cataacaccc	360
tgcatgagct gtttagcctg gttaaagtcc ttgatgaatt tattgacttc gacgaatgta	420
tttcagaaac cagcagcaat acgacggcga cggcttggat ttaacaaatc tgggttttca	480
cgctcttcag gtgtcatcga agacacaatg gcacgtttac gagcaatctg gcgttcatcc	540
accttcatgt tttga	555

&lt;210&gt; 107

&lt;211&gt; 396

&lt;212&gt; DNA

&lt;213&gt; Streptococcus pneumoniae

&lt;400&gt; 107

ttatcaggat caaaatatca aatgaaaaa gaacaatttt atccgctagg aatttttcta	60
gctgctatgt tggcgaggact tgtccgatat ctagtttcca cctggttacc agccagtcca	120
gatttccctt ggggcactct ctttgtcaac tatctgggaa tttctgctt gatttatctt	180
gtcaagggct atctggtcta taaggggact agtaagggct tgattttagc actggggagc	240
ggtttttgcg gaggtttaac aactttttcc agcctaagtc ttgatactgt gaagctgctt	300
gatacagggc gttatcttag tttgatactg tatttgcctt tgagtatcgg tggaggcctg	360
cttttagctt actatttggg gaggaagaaa tggtaa	396

&lt;210&gt; 108

&lt;211&gt; 1998

&lt;212&gt; DNA

&lt;213&gt; Streptococcus pneumoniae

&lt;400&gt; 108

aaaaatatgg coattagtca gatgaaaaga atctctctac tattttctaa aagtagtctt	60
gatgatgttt taaaactat tcaagaacta gtagcagtgc agttccgtga tttaaagggt	120
caggataact ggtcagaagc tctagaaaaa gatgaagttg tatttccaac tattcaaatt	180
tttcatactt ctaattccaa tcatgggggtt attgagggaa atgatgcctt gaactatttg	240
atgaatcaac aacaacattt agaagcaact gtatagaaat tacaagaata cctaccgaaa	300
gaaaacacgt ttaaatattt gcagcaacct cagataacta cctcttatga agaattagag	360
aaatttggta aagctaattg tgctgagggt gttcttaaaa aagtgaatca tcaaatatc	420

agagttcatg aattagaaag acacattcaa agtaataatg aggaaataga gcgattaata	480
aagtgggaaa aattagaaat tgttctctgcg aatttagaac aattttcttt ctgtaaagga	540
aaagtcggaa caattccaag gactgaagat aatcgcttat acaatagtct tttagaaaac	600
aattttgaag ttcaagaat attttctaata gatagagagt acggtgttgt tgtttcttat	660
cagtcctagt actctataga ttttgatgaa tacttatttg aaccatttga ttattctaga	720
aagggaattac cgaagcagcg agtagtagat ttagatcaag aaaacatgca gtttaataact	780
gaaaaagaga atattatcgc atcgttgcaa gattcaaaga aatatttgat agatttaca	840
tggcaaatag actatatttt atctatctat gctcgtcaaa tctctaagaa taactttttg	900
tgcactccgc atctagtgtc attagaagga tggatagaag aaactcgtat tttatatttt	960
ataaaagtta tggatgagca ttttggacat tctatttata ttatgaatc ggaacattg	1020
acggataatc aagatgaaat acctatcaa ttaacgaatc attctttaat tgaacattt	1080
gaattattga cagaatgta tgctctgcc aaatattatg agaaagatcc tacacctgta	1140
ttagcaccat tttactttac attttttgga atgatgggtg ctgatttagg ctatggttta	1200
ctattgtttt taggaacaat gttagcatta aaaatttttc atctaccttc agcaactaag	1260
agatttttaa aattctttta tatattaggg gtacccgttg caatttggg tggaaactat	1320
ggctcatttt ttgcatatga gttgccatt catctgatat ctacaacctc tgatgtcatg	1380
actatattag tagtgctcag tgtgtttggg ttattacag tatttgcagg tttgttagct	1440
tcaggactac aaaaagtaag aatgaataaa tatgcagaag catataattc aggatttgcg	1500
tgggtgtgta ttctgcttg cttgttattt attgctgttg gaatgttgat gctgatatg	1560
agaccgttat ttgtattagg gaaatgggta tctattttta atgctgtggg gattttgatt	1620
gtttctatta ttcaaacc aaagcttgta ggtattggag caggattgtt taatctatat	1680
aacatttc attatatagg tgatttagtt agtttcaact gattgatggc attaggatta	1740
tctggagcaa gtatagcatt agctttcaat ttaattgttg gtttgttcc gggaataattg	1800
gctaaactga caattggatt agtattattc attcttttac atgcgatcaa tatttttcta	1860
tcgttactat caggatatgt tcatggagca cgtctgatat ttgttggaatt ttttggttaag	1920
ttttatgagg gtggaggaaa accatttcaa cctttgaagg ctctgagaa atatattaag	1980
gttattacaa agaattaa	1998

<210> 109  
 <211> 915  
 <212> DNA  
 <213> Streptococcus pneumoniae  
  
 <400> 109  
 gatcaaaaat gtgggagtggt tgaatgaag attataggtg tcatattgg cggacaaca 60  
 attaaggcag atttatcga tgagtttga acgagtttga atcatttcaa agagatagaa 120  
 acaattattg actatgattt gggaacgaat cagatattaa atcaggctcg tgaatttaatt 180  
 ggtgagtata ctttaaatca ttcaattgat ggtgttggga ttccactgc tggagttggt 240  
 aatgctaata ctggagaaat catctatgca ggctatacaa taccagggtg tatcggagta 300  
 aactttactg cgaataaga aaaacgtttt gggttgata cttttgttga aaatgatgtt 360  
 aattgtgctg cattaggtga attgtggaag ggacaagcca aagataagaa aaatgtagta 420  
 atggttacta ttggaacagg tataggaggc agtattattg tcaacggaca aattgttaac 480  
 ggatttaact atactgctgg tgaagtaggt tatattcctg taggtaattc ggattggcaa 540  
 agtaagcct caacaaccgc attgattcat ttatatcaaa aaaagagctt gaaaactaat 600  
 caaactggac gtactttctt cactgattta agatctggag ataaagttgc tgaagaaact 660  
 ttgaaattt ttgtagaaaa tctaacaaa ggtttattaa cgatttctta tctacttaat 720  
 ccagaaattc tcatattagg aggtgggatt ctggatagta aggatatttt gttacctgaa 780  
 attcaaagtt ctttagctaa aaatgcaatg gataataggt tttacctaa aaatcttgtg 840  
 gcagctacat taggaaatga agctggctgt ataggagctg taaaaaattt cttagataga 900  
 atttctaata aatag 915

<210> 110  
 <211> 930  
 <212> DNA  
 <213> Streptococcus pneumoniae  
  
 <400> 110  
 aggagaaatc tgatgaaga tttaactaaa tacaaaggcg ttatccctgc attttatgct 60  
 tgctatgatg aaaatggtga aattagccaa gatcgtgtaa aatctctggt acaattattc 120  
 attgacaaag gtgtaaaagg tatctatgta aatggttctt cagggtgaatg tatttaccaa 180  
 agtgtagaag atcgtaacaa aattattgaa gctgttatgg aagttgctaa aggtaaatta 240  
 acagttatca accatattgc atgtaataac acgaaagata gtatcgaatt ggcaaaacat 300



tcagaaagtg ttggagtcga tgctattgca gctatccac ctatttattt caaattgcc	360
gagtatccaa tcgcagcata ttggaatgca atgagtgaag ctgcgtcaaa tacagatttt	420
attatctata atattccaca attggcaggg gttgcgttga ctggtagttt gtatgcaaca	480
atgcgtcaaa atcctcgtgt gattggagtt aaaaattctt ctatgcctgt acaagatatt	540
caaatgtttg tagctgcagg tggagaagat tacattgtat tcaatggctc agatgaacaa	600
tatcttggtg gtcgcttgat gggagcagaa gctggtattg gtggtactta tggcgttatg	660
ccagatttgt tcttgaaatt ggaagtttg attcaagaac gagatttaga tacagctaaa	720
aaacttcaat atgctatcaa tgaagttatc tataagatga tatcaggtaa ggcaaatatg	780
tatgctgtag caaaagaagt ttgctgtcta aatgaaaaac ttgatttagg ttctgtctgt	840
caacctttag aagcattggc agaaggtgac ttggaagttg caaaacaagc agcagaactt	900
attcaacaag cagcaaaaga atttttataa	930

<210> 111  
 <211> 759  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 111 gtgattggag gcaagaatat ggataaagat tatatatataa aagtgaaggg gctgtatcat	60
caatttttac taggaaataa taaaacgttg caagtgtga aaaatgtttc tctttctgct	120
tcgagaggag aatttataag tattctagga attagtgtgt ctggaaagtc aactttatta	180
aaatgtatct ctagtgtgtc tgaaccaaca agtggggaag taattttaaa tggaaatcaac	240
ccctataaaa tcagaaatgc aaaattgtca agtataagac gtaacgaagt atcttttata	300
tttcaagcat acaatttaac acctccctg cgggtaatag aaaatatagc acttcccttg	360
cgattatcac aaaaaaatt aactattaaa aatgtagaaa acttactcaa aagaatgaag	420
tttaattgctg gcttaaacga ttttgttga actctgtctg gtggagaaca acagaaggtt	480
gctatagcta gagcgggtat tgctgatagt gatataatat ttgctgatga gccaaactggg	540
gctttagaca gcgtttctcg tgaagtaatt tttgaattat tgagagagtt agtaggggcg	600
ggtaagtgtg taattatggt aacgcacgat atagaattgg cctcgaaaac tgatcgtgca	660
tttaatttga aagatggaaa aattttcaaa gaacttcata gacctagcgg ggaagagttg	720
tataaaatct tagagggtaca atcaactacg gaggaatag	759

<210> 112  
 <211> 1611  
 <212> DNA  
 <213> *Streptococcus pneumoniae*

<400> 112  
 cttatgaatt atttaaaatt tataaaaaag actaaactta tattaatggg gatttttaata 60  
 ttctatctt cgtttaatgg tgccttgctc tcaggaatta ttgtatatgc aggtagttta 120  
 aatcagactt cctcatttct tgacgttcta agatttgggtg ctataagtat tctgggttgg 180  
 tcagcaatct atatatcaaa ttactattta gaagtaacgg aagcatcaat aactaaagat 240  
 ataaatgtaa aaattaaaca aggatatttt agagaacagt atctttcttc tgaatgggtt 300  
 aaagattact catctattat ttcagttttg tcaaatgatt tgagattaat cgaagaaaac 360  
 ttttttagac aaatttttga aataatttct tcaatattgt tgttcacgt ctctctaagt 420  
 tttatgcttt atttaaaact cctagtttca ataatttta ttgtattatc agcattgocg 480  
 attatagtag ctgtctttat gaagaagatg ttgtctaatt cagctaataa gtaacttaac 540  
 agcaatgcag agtatactca cataataaaa gaaattttta atgggttttaa gacattaaaa 600  
 tcttattctg ttactaaaga aataattagt ttgtcggata aaaagttgga taaactagaa 660  
 gattctactt ttaatttgaa acgatcagag gttctttcaa aattgggtgc agtactaatt 720  
 tcagggtttt gctttctagt tcccttggtt gtaggatgtt attttgtaat ttatcataaa 780  
 agcctttctt ttagtgaact gataggtatt ttcttagcaa acgataaagt tttagggccg 840  
 atacaatcaa ttgcctattc attaaaaag ataaatacaa ccaaagattt aaggaaacgg 900  
 tttttaaat acttaagtgg agagaagaat tttatagacg ctgaacatga taataacgga 960  
 ctgtatactt catcaataga tgagatacac atgaagatg ttgtatatc tattacacca 1020  
 gaaaataaat taagtattga ctctctcatt aagtcaccat ttaggggtatt attaacagga 1080  
 actctcggtg gtgggaaac aacgatatta aatttaatta atgggtcttt aaagccacaa 1140  
 aaaggttatg taaatttggt atcacatggg aaaaagagtt cagattcaat accaacagtt 1200  
 gatcagacac catatatttt tgacactact attcgtgaga acgtaacctt atttcaaaat 1260  
 gaatattttt cagatgatca gataattgag gtgttaaaaa aggttaactc atatgaagaa 1320  
 ttagaaaaga tagatatact aaattatcaa tgggtgaaa atggtagtaa tttgtctgga 1380  
 ggtcaaaaac aaaaaatagc tttagctaga gctctgatta gaaataataa agtgacttta 1440

tttgacgaaa tatcagctaa tttagataat gataattcaa attccatata tgatattctg 1500  
 ttcaatttag gtatttcatt tattgaagtt tcacatcatt atgacttaaa tgacaagaga 1560  
 tacactgata tatataaatt ggaaaaatgga acacttttca aaattaagtg a 1611

<210> 113  
 <211> 1953  
 <212> DNA  
 <213> *Streptococcus pneumoniae*

<400> 113  
 ggggtcatcac ttgtgaatca gagtaatgaa atatttttaa atacaatata ggaaaaaaca 60  
 cactataaga ggtctatctc ttcacatctt caggaattaa aaaaatggaa ttctttgata 120  
 tttaggagag ggcgatattg gtaccaattt gataattgga gtatacccaa tactgggtgta 180  
 ataaaggaat ggtatgtcga agataatgac tggattoccta aggttagttc ttocagttta 240  
 gacaattcaa attcacttaa aattgtcttt gatgtggaaa aatatacttt gccagatgat 300  
 agtatcttac aggatgttag gatttcatct gatgagagtt atatagtttt ggtgtgatcc 360  
 tcactgcgta cgactaattt aattggaata aaaaaatata ctatggaaga gttatttctc 420  
 attgaagata tttcagtaga aagtctgttc tatttgggta aattcgggtgt tatgtatact 480  
 cgctccaag agtacgggag gccaaagtaa ttattctata aatcttttga tagctttaca 540  
 gaagaggaac tgtttgaaga aaacgaatgc tcttttcgat taaaaatagt ccacattgat 600  
 agcaataatt gttttgtaaa gtcagtagac ttccagaaag gcaggatttt tctatacagt 660  
 ttcgaccgga cgggttttgt tagacactct tacacagaaa cggtagcccc aacacctaga 720  
 gatatcgctc tatttagtac agaaaaagca cagtactttc ttgggctctc ttcaactgaa 780  
 gaaaagaag atcaactaat tctaagggaa acttcttcgg gagacagagt ttctattact 840  
 ataccatata aagatagggc tagaagagtt cactgtatag gacgttatat ccttttagac 900  
 tgttcaaagc cagcggaatto cgtcttttat cttgctacat ttaaaaaata tagtttgagg 960  
 gaggttgagtg ttaataagat aacatttgat gaaaagacaa cactttatga aaattcattt 1020  
 tctgatcgtg ttctctctgtt tttagagagg agaacttttt acgaaaaaat acttagtttt 1080  
 gatttaaatag aaaatacatt aaaaactgaa tttgaaagac ccatcataaa atcgataaat 1140  
 actaaatatt tttctaaggt gatttggact aaaaaagaca gttagatgtg aagcataccg 1200  
 atttctttat tttgaaaag tgaagatacg gatgagctcc caagaagaaa aaatgtatt 1260

ctaagtgtct atggggccta tggtaagaat gataattctg atttagatga aattatgtta 1320  
 tctataatag atgcaggttt tatttatgct atagttcatg taagaggtgg aggcattttg 1380  
 gglggtgaat ggtatcgctc aggaaggcg ttaacaaat ggaattctat cagagatttt 1440  
 attgagggag taaattattt aagggaanaat gatgtaattg acagtaagcg attaggttta 1500  
 ataacttcta gtgcagggtg aataattgct ggtgcggtgt tgaacgagga aaaaaattta 1560  
 ttgcaagca ttctcttatt ttgcacctt ataatcctt atgatacact tcagaatcca 1620  
 aatgatcctc ttcttaagac cgaaatagca gagtggggag atattagaga cctgaagta 1680  
 aaggcttata taaagtcata ttgcctatg caaaatattg aaaaggcagc agactcaaat 1740  
 actgttatag ttaatatattt ggttgagaaa gaccatata ttaataataa cgaagtata 1800  
 gagtgtgcaa agaaattaaa ttctattgga gttaaaagt tttgtattt aaacaaagca 1860  
 gctggtcatg gaggttttac tccatcagat gttctcttaa tgattgatac tttaaattat 1920  
 ttctttgaag aagtgggaag gaataactta tga 1953

<210> 114

<211> 1449

<212> DNA

<213> Streptococcus pneumoniae

<400> 114

aacgggggtg accaagtgat tgatgggaaa cgattattat ttagtttgac catagtcagt 60  
 tatgccttga cgctagtaag tggattgtg tatctgttta ataataataa tgtagctta 120  
 ctttctactt tattgttctt gttggttagt agcttaattg cttgttgga tgatatacag 180  
 tattacttaa tccattttat ttctatttta accatttttg tattctggt atcaagaccg 240  
 accattgatt attttaggga tgggtgcttg gataacctat atccaatagc ctatcgtttt 300  
 gcctttatag ttgtcatgat ttgcattotg ggcttgacca caggaggcat tctggctcgt 360  
 tacttcatag ctagggaaga aataaaagta gcaaatatag gaaattctct aaaagaggtt 420  
 tatatcaagc ggttacgctt tgtatcacta ggagtttttc ttctaactta tcctttctat 480  
 ttcatcgggt tatttgaacg gctcttgat cgtttgcaga ctctacta tgcctactat 540  
 gcaaattttg aaagtaaac gcctattttt acctacattt tgtctacctt tacggtctat 600  
 gcaatgtgta tgatctggc aaccaagcca aagaaattgc aggccacagc agtgcctgtc 660  
 tcctttattg cagctaatac tattcatttg gcaattggga cacgaaatcc ctttatttta 720

```

agtatatttat ttgcttttgt ttattacttt atgcggggagc aaactgaaaa aggaaaaatgg      780
attgggttta aagaaaagtt agcgattttt gtaggttctc ctattctcat gttagcgatg      840
ggagtactca attatgtacg ggataatgtc caagtttccc atacagggtt ctgggatatc      900
ttacttgact ttatctataa acaagggact agttttgggt ttctggctcg aggttttcta      960
tttaacagta gctccctta cggagatttc cgtaatttta cttttggtec tgtctctgat      1020
tattttgcaa gggggagttt gggagccatt ttccggaggaa aagcctttga acatacaacc      1080
aatagtgtgg aactagctat tgatagtaat agttatgccc acaatctatc ctatcttgte      1140
ttgaacaagg aatacttgaa agggcatggg atcgggaagta gttatatcat ggagttgtat      1200
accgactatg gtatgattgg agtctttctg cttagtttct tactcggcgt attatttata      1260
gccatgctgc aagtagccta tcgctcaagg acaatcctat ttgctttatc cctactcate      1320
ttgaataatc tattctttat gccagaagc agcttttcag aaagtttctt caatttattt      1380
acaatgcaat tctggggaat tgttcttggt attatatttg tagcaaaaat gcttacaaag      1440
gaaaactag                                     1449

```

```

<210> 115
<211> 831
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 115
tcaatgcaat cgaggaggaa agagatgaag aaaacaagct ctaaactctt tgtagtacc      60
tacctgcttt ggattgcgct ctttgtattg gcacccttgg tcttgatttt cggccaatcc      120
tttttcaaca tcgaagccca gttcagttta gaaaattaca aatcttactt tgcgtcacaa      180
aacttgacct atcttaaaat gagtttcaac tcagtgtctt atgcaggcat tgtgaacctt      240
gtggcactgc ttatcagtta tccgacggcc ctctttttga cccgtctcaa acacgctcaa      300
ctctggctca tgctgattat ccttcctacc tggatcaatt tgctccttaa agcctatgct      360
tttatcggga tttttgggtc aaatggctct attaaccaat tcttggaatt tatcgggaatt      420
ggttcaaac agttgctttt taccgatttt tcttttatct ttgtcgcaag ctacatcgag      480
ctccctttta tgattttgcc gattttcaat gtcttagacg atatggataa taatctcatc      540
aatgctagtt atgaccttgg tgcaactaag tgggagacct tccgtcatgt catcttccct      600
ctatctatga acggtgtgcg aagtgggggt cagtcggtct ttatcccaag tttgagttct      660

```

ttcatgctga cccgtttgat tgggtgggaac cgcgttatca ccttggggac ggctattgag 720  
 cagaattttc taaccaatga caactatggg atgggttcaa ccatcggtgt gattctcacc 780  
 ctgaccatgt tcatcaccat gtgggtgact aaggaaagga gagaacgatg a 831

<210> 116  
 <211> 771  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 116  
 aaaggaacc gtatgacaga tgcgatttta caggatcag acctgtccgt ttattataat 60  
 aaaaagaagg ctttgaatag tgtttcccta tctttccaac ctaaggaaat tacagccttg 120  
 attggtccat ctggatcagg gaagtcaccc ctctcaagt ctctcaaccg catgggagat 180  
 ctcaatccag aggtgaccac aactggatcc gtggtgtaca atggtcacaa catctacagt 240  
 ccgctacag atacggttga attacgtaag gaaatcggaa tggttttcca acaacctaat 300  
 cctttcccta tgactatcta tgagaatgt gtctacgggc ttctgtatcaa tggaattaag 360  
 gataagcagg ttctggatga agccgtagaa aaagccttgc aaggtgcctc tatctgggat 420  
 gaggtcaagg atcgtctata tgattcagct attggattgt cagggtgtca acagcagcgt 480  
 gtctgcgtgg cccgtgtctt ggcaactagt cctaaaatca tctcttggga tgagccaact 540  
 tcggtcttgg atccgatctc agctggtaaa attgaggaag ccttgtatgg tctaaaagac 600  
 aagtacacca tgcttctggt aaccggttcc atgcagcaag ctacacgtat ctctgataag 660  
 acaggatttt tctatagatg agatttgatt gaatttaatg ataccaagca gatgttcctt 720  
 gatcccaaac acaaggaaac ggaagactat attacaggaa aatttgata a 771

<210> 117  
 <211> 912  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 117  
 ttacgaaga aagaggaaag aaaaattatg cgcgcctaaga aattagataa acttgcaaca 60  
 gctgtcctct atacgattgc tagcatcatt gtgacaatct tggcttcctt gattctctat 120  
 atcttggttc ggggcttgcc ccatatctct tgggtcttct tgactggaag gtctcttctg 180  
 tttcaagcag gtggtgggat tggcattcag ctttacaatt ccttttctct attggtcatt 240

```

accttgatta tttctgtacc tctttctatg ggagctggga tttacttggc tgaatatgct 300
aaaaaaggtc ctgttaccac ctttgtgcgg acttgttattg aaattttgtc ctctttacca 360
tcagtggttg .tgggtctctt tggttacttg atctttgtag tccagtttga gtagtgattt 420
tcaatcattt caggtgcctt ggocctgaca gtctttaact tgcctcagat gacgcgtaat 480
gtagaggata gtttgaacac cgttcaccat acccaacgtg aggcgtgtct ggctcttggg 540
attttctcgt gggagacagt ggttcagtgt gttattccgg aagcgcttcc aggtatttga 600
acgggtgtcg tcttggcacc tggctgtatc tttggcgaag ctgcagctct gatctataca 660
gcagggcaat cggcgccagc tcttgactgg tctaactgga atatcctcag tgtgactagc 720
cccatctcta tcttcctgca agcagaaacc ttggctgtcc atatctggaa agtcaaatagt 780
gaaggcacta ttccagatgg aaccattgta tcagcaggtt ctgcgcgtgt gctcctgac 840
tttatcctga tttttaactt tggagctcgt aagttcgaa gctatctaca caagaaatta 900
accgctgcct aa 912

```

&lt;210&gt; 118

&lt;211&gt; 1800

&lt;212&gt; DNA

<213> *Streptococcus pneumoniae*

&lt;400&gt; 118

```

gcgaatttat atctaaaagg gatattaaag aaaggagata tgcttatgaa gatttacaaa 60
aaactatttg cttatgtcca agataagaaa tatcttgggg ttttggccat aattttttct 120
gctatatctg ctgcacttac agtatatgga tattatttaa tctacaaatt tctagataag 180
ttaataatta attcaaaact atccgggtgca gagagtatag cattaaaaac tgttattaca 240
ctaacaagtg gagcgatatt ttattttgtc tcaggaaatg tttcacatat cttgggattc 300
aggcttgaaa caaattttaag aaaaagggga atcgatggtc tggaaaaagc aagttttagg 360
ttctttgact taaatccatc tgggtcaata agaaagatta tagatgacaa tgctgcacaa 420
actcatcagg tggtagcaca catgattccc gatagtcttc aggcaataat cacaccgta 480
cttgacttgg cacttggcct tatagtaagt ataagagttg gcataatttt gcttgcctct 540
actataattg gtggcttaat tttaggggca atgatgggag agcaagaatt tatgaagata 600
taccaagaat ccctatctaa actaagtgtc gaaactgttg agtacgtgag aggaatgcaa 660
gttgtaaaaa tatttaaagc aaatgtagag tcttttaaaa gcttttataa ggcgataaaa 720

```

```

gattactcaa agtatgctta tgattattcc ctatcttgta aaaggcotta tgttttgtat      780
caatgggttat tttttggact gattgcaatt ttaattatto ctatagttta ttttatgact      840
agcttagcta gcgcaaaggt gatttttactt gagcttatca tgattttatt tttatcagga      900
gttctctttg tttcattcat gagaatgatg tggtagtcca tgtatatttc tcaaggaaat      960
tatgcagtag atactttaga ggcgctttac gaagatatgc aaaaagacaa attagtgcatt     1020
ggtaatgtca ataattttaa aaactataat atagaatttg agaattgttag ctttgcttat     1080
aatgataaag ctgtcattga aaatttatcc ttttaattag aagaaggaaa gtcctacgca     1140
cttgctgggt catctggatc aggcaaatca acagtagcaa aacttatatc aggtttttac     1200
aatgttaata aaggagcat aaagataggg gggatagcaa taagtgaata ttctgacgaa     1260
gccttaatta aagccatttc ctttgttttt caagattcaa aattattcaa gaagagcatt     1320
tatgataatg tagcggttag taataaagat gcgacgaaag atgacgttat gagagcotta     1380
aaattagcag gatgcgattt aatatttagc aaattcccag aaagagaaaa tacaatcata     1440
ggctcaaaag gtgttttatt atccgggtga gaaaaacaaa gaattgcaat tgctagagca     1500
attttaaagg attccaaaat tattattatg gatgaagcat cagcatctat tgaccagat     1560
aacgagtttg aattgcaaaa agctttttaa aatcttatga aggataaaac agttatcatg     1620
attgcacaca ggctatctac aattaaagac cttgatgaaa ttattgtcat ggatagtgga     1680
aaaattatag aaagagggtc tgacaaagaa ttaatgtcaa aagatacaag gtataagagc     1740
ctgcaagaga tgtttaacag tgccaatgaa tggaggggtt caaatgaaag agttttataa     1800

<210> 119
<211> 1791
<212> DNA
<213> Streptococcus pneumoniae

<400> 119
tgcaaaaga tacaaggat aagagcctgc aagagatgtt taacagtgcg aatgaatgga      60
gggtttcaaa tgaagaggtt ttataaaaa agattttgctc ttacagatgg aggagcaaga     120
aatttaagta aagcaacact ggcttcattt ttctgtttatt gtataaacat gcttcctgcc     180
atattactta tgatttttgc tcaggaaagt ttggaaaata tgggcaaaag caatggcttt     240
tatatagtat tctcagtttt gatthttgata gcaatgtata ttttgctttc tatogaatc     300
gataaattat ataacacaac ctatcaagaa agtgcagatt taagaataag gacagcggag     360

```



```

aatttatcaa aattacctct atcttacttt tctaaacatg acatttcga catttcacaa 420
acaatcatgg ctgatattga aggcataag catgcaatga gccactcaat accaaaggtg 480
ggcgccatgg tactgttttt ccattaata tctgtaatga tgctagcggg caatgtcaag 540
atgggttttag ctgtaattat tccatctatt ttaagcttta tatttatacc tttatctaaa 600
aaatcagcag ttaattggaca gaatagatat tatgatgtct taagaaaaaa ctcagaaagc 660
tttcaagaaa atatcgaaat gcaaatggag attaaagcat ataatttato gaaggatatt 720
aaagatgact tatataaaaa aatggaagat agtgagaaag tacacttaaa ggcggaagta 780
actacaattt taactttgtc tatacttca atatttagct ttatatctct tgctgttgtg 840
atatttgtcg gcgtaaatct aattattaat aaagagataa attctctcta ccttatagga 900
tatttactag ctgctatgaa gataaaagac tctttagatg catctaaaga ggccttgatg 960
gaaatatttt atttatcgcc caaaatagaa agattaaaag aaattcaaaa tcaagattta 1020
caagaaggcg atgactatag cttaaaaaaa ttgatattg atctaaaaga tggttagttt 1080
gcctacaata aagacgcгаа agttttaaat ggtgtaagtt ttaaagctaa gcaggagag 1140
gtcactgctt tggtaggtgc aagtggtgc ggtaaaaaa ctatcttgaa acttatatca 1200
agactttatg attatgacaa gggacaaatc ttaatcgatg gcaaagatat aaaggaata 1260
tcaacagaat ccccttttga taagggtgtc attgttttcc aagatgtggt tctctttaat 1320
caaagcggtt tggaaaatat tagaatcggg aagcaagatg caagtgcga agaggttaaa 1380
agagcagcaa aacttgcaaa ttgcacagat tttatagaaa aaatggataa aggtttcgat 1440
acagtatttg gtgaaaacgg agctgagcta tcaggaggag aaagacaaag attatcaata 1500
gccagagcct tcttaaaaga tgcgccgata ttgatcttag atgagataac agcaagcctt 1560
gatgttaaca acgagaaaaa gattcaagag tctttaataa atttagttaa agataaaact 1620
gttghtaatca ttccacatag aatgaaatcc atagaaaatg cagacaagat agtagttctt 1680
caaaacggaa gagtagaag cgaaggtaag catgaagagc ttttcaaaa atcaaaaatt 1740
tacaaaaatt taatagaaaa gacaaaaatg gcagaagaat ttatttatta g 1791

```

&lt;210&gt; 120

&lt;211&gt; 675

&lt;212&gt; DNA

<213> *Streptococcus pneumoniae*

&lt;400&gt; 120

```

aaaagacaaa aatggcagaa gaattttattt attaggagga ctacaatgga taataaaaaa    60
ttaaagtaa aagatttagt aagcatcggt gtttttggcg taattttattt tgccttcacg    120
tttggagttg gtatgatggg cttgattcca atattgttct taatataccc gacagtatta    180
gccatagttg caggaactgt tgttatgtta tttatggcta aggttcaaaa gccatgggca    240
ctattttata ttggtatgat atcaccactt gtgatgtttg cagctggcta tacctacgta    300
gttgtggttt tatcacttat agtaatgata atagcagaat taattagaaa gattggtaat    360
tataatcat ttaaatacaa tatgctttct tatgcaatct tcagcacatg gatatgtagc    420
tctttaatgc aaatgctttt agcaaaagaa aaatatatgg agtggctctt gatgactatg    480
ggaaaagatt atgttgatgt attagaaaag ttaataactt atcctcacat ggcttttagta    540
gccttaggtg ctttcttagg aggaattctt ggagcatata taggcaaggc tctattgaaa    600
aaacactttt caaatggatt atattgtgtg ggatacttta ctccttgccct aattttatgg    660
tgctatctga attaa                                     675

```

&lt;210&gt; 121

&lt;211&gt; 636

&lt;212&gt; DNA

<213> *Streptococcus pneumoniae*

&lt;400&gt; 121

```

tgtattagaa aagttaataa cttatcctca catggcttta gtagccttag gtgctttctt    60
aggaggaatt cttggagcat atataggcaa ggctctattg aaaaaacact tttcaaatgg    120
attatattgt gtgggatact ttactccttg cctaatttta tgggtgctatc tgaattaaac    180
cctatagtta agatgttttt gagtatacct attgttatta gaatgtttat tttaccattt    240
atggcagcaa gctttatgat aaagacctcg gatgtaggcg caataatttc atcgatggat    300
aagcttaaga tttcaagaa tgtatccata cctattgcgg ttatgttttag attcttccca    360
tcttttaagg aggagaagaa aaacatcaaa atggctatga gagtaagagg gataaatttt    420
aaaaaccacg tcaaatatct tgaatatgtt tctgtgccac tactcattat atcatctaatt    480
atatcagatg acattgcaaa agcggcagaa acaaggcaa tagaaaatcc aattgcacag    540
accagatata ttgcgtaaaa gatacagcta attgattttg tttatgtttt agcgggttgc    600
ggacttattg tgggaggcct aatatggttg aaataa                                     636

```

&lt;210&gt; 122

<211> 1173  
 <212> DNA  
 <213> *Streptococcus pneumoniae*  
  
 <400> 122  
 ttttggtttat gtttttagcgg ttgctggact tattgtggga ggcttaatat gggtgaataa 60  
 aaaaatttaa gtcttgatta tggggaagag catatattag atgatataatc actatccata 120  
 gccgaggggag agtcgctgct atttacagga aaaagtggaa atggtaagtc atctttaata 180  
 aattcaatca atggactagc tgtaaggat gataacgcaa agacaaggg cgaaataatt 240  
 attgatggta agaataaaa aaatttggaa ctttatcaa tctcaatgct tgtttcaact 300  
 gtttttcaaa atcctaagac atattttttt aatgtcaata cgacattaga attattattt 360  
 tatttgaaa atatcggtct tgcaagagaa gagatggaca ggcgtttgaa ggatatactt 420  
 gagatattcc cgataaaaaa tcttttgac agaaatatat ttaatctatc cggcggtgaa 480  
 aaacaaattc tttgcattgc agcttcttat atagcaggta caaagattat agttatggat 540  
 gagccttcac cgaatttaga tattaaaagc ataagtgttt tggcaaagat gctaaagata 600  
 ttaaagaga aaggcataag cataattgtt gcagagcata gaatttatta ttgatggac 660  
 atagttgacc gtgtattttt aatagataaa ggaaagccta aaaaactta tactagaagt 720  
 gaatttttaa agctagataa aaatgaatta aatgctttaa gttaagaga taagaatta 780  
 agtaaatata aagtctctta tttaaagaa ggtggagagt atcagataaa aaatcttagt 840  
 tacaattta ctgatgatga gtgtttaagc ttaaagata ttctgttcaa gcttgggaaa 900  
 atttatggca taataggatc caacggacga ggaaatcaa cgcttttaag atgtttaata 960  
 ggtcttgaga aaaaatcaaa agaagaaatt tattttaagg gagagaagct atctaaaaaa 1020  
 gaaagactca aaactcttc acttggtatg caagatgtaa atcatcaatt attcacagat 1080  
 gaagtattca acgagcttag attaggagta aagaattttg atgaagaaa ggcgaaatc 1140  
 attttaaacc ccaattattc accccaatc taa 1173

<210> 123  
 <211> 276  
 <212> DNA  
 <213> *Streptococcus pneumoniae*  
  
 <400> 123  
 tttgggttaa aagatttatg cctggacgaa ttattgaaa ggcacccgat gagtttatca 60  
 ggaggggcaa agcaaggct tgcaatagca tctgttatgt gcaagaattc tccatttgct 120

ttttttgacg aaccttcaag tggatggat tattccaata tgataaaat atctgaactg 180  
 attaataagt ataaaacat ggataaaata atttttattg ttcccatga tatagaattt 240  
 ttaaatgaag tggcagatga aatttttgaa ttgtaa 276

<210> 124  
 <211> 975  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 124  
 aaaggacgag agagctcaat ggatattaga ccgcaatcaa gtgatgaact tgattcgcaa 60  
 agaagagtaa ggagagacat gtcaaatagt ttaaaagggg ctttactaac agttgtggct 120  
 ggtattgctt ggggggtgtc aggaacgagt ggccaatacc taatggcaca cggaatttcg 180  
 gctctggtct tgactaactt gcgtctttta atcgctgggt gaattctcat gctcttggct 240  
 tatgtctacty caaaggataa aatactggtc tttttaagg atagaaagag ttgtctgtct 300  
 cttcttattt ttgctctgat tggcttttt ctcaaccaat tcgcctatct gtctgtctatt 360  
 caggagacca atgcgggaac agcgacgggt cttcagtatg ttgtctctgt cggaatttta 420  
 atttatagct gtatcaagga taggggtggca ccgacactgg gagagatagt ttccatcata 480  
 ttgcgccatg gaggaacctt cctgatcgca acacatgggc agttggacca gttatcoatg 540  
 acacctgctg gtctgtttctg gggctctttt totgccttga cttatgctct gtatatoatt 600  
 ttacccatag ccttgattaa aaagtggggg agcagcttgg tcattgggtg gggaatggtc 660  
 atagcagggt tggtcgccct tccttttaca ggggtttctac aggccgatat cccgactagt 720  
 ctgtattttc tccttcgctt tgcaggcatt atccttatcg ggactgtctt tgctatataca 780  
 gotttcttta aaggagccag tctgatagga cgggtcaagt caagcttggt ggcttcaatt 840  
 gagccaatat cggcgatttt ctttgccctt ttaataatga atgaacaatt ttatcccat 900  
 gattttcttg gtatggcaat gatattgttt gctgtaactt tgatttcttt gaaagattta 960  
 ttcttagaaa aataa 975

<210> 125  
 <211> 366  
 <212> DNA  
 <213> Streptococcus pneumoniae  
 <400> 125

```

atatctcaaa acgcattatc gctgttttgg tacctaatat tgttgaagaa ggcgaaactc      60
cacaggaagc ctacgatttg gaagccatta tgtacaatcc aaaaatcgtc tctcactctg      120
ttcaagatgc tgctcttggc gaaggagaag gttgcctgtc tgttgaccgt aacgtgcctg      180
gctatgttgt tcgccatgcc cgcgttactg ttgactactt tgacaaagat ggagaaaaaac      240
accgtatcaa actcaaaaggc tacaactcca ttgttgttca gcatgaaatt gaccacatta      300
acggtatcat gttttacgat cgcacatcat aaaaagaccg atttgcagtt aaagatggtt      360
tactga                                          366

```

&lt;210&gt; 126

&lt;211&gt; 261

&lt;212&gt; DNA

<213> *Streptococcus pneumoniae*

&lt;400&gt; 126

```

gtttactctt taanaaagat agaaatgaga gaaatcatgc tactgcaatt attttcttta      60
tatttcgaga gtttgatctt gaccaccatc cttgttctga ttttttaggg gatttggatt      120
ggtctgagag ccatgtcggg agttgataag acagccaggg ctgcacaagg ccatctctat      180
gatagtatta tgattggagt cttggttgtc ccagtattat cctttgcggt tatgagttta      240
attcttgttt tcaaggcata a                                          261

```

&lt;210&gt; 127

&lt;211&gt; 579

&lt;212&gt; DNA

<213> *Streptococcus pneumoniae*

&lt;400&gt; 127

```

gatgccagta gatggcgac gcttggccta tcaaaaatta aagaataat gcaaagaag      60
tatgtaaaaa tcctctactc ctccaccaatt ggtattctat cactgttagc tgatgacct      120
tatttgtatg gaatttgggt tcaggagcag aagcattttg agaggggact aggagatgaa      180
acgatagaag aagttgttag tcactctatt ttagaccagc ttattgcttg cttagatgat      240
tactttaaag gcaagcctca ggatttatcc aacttgcctc tggcgccaat cggaaacgaat      300
tttgaaaaaag gatttggga ctatttacag ggcattcctt atggtcagac agtgacctat      360
ggacaatttg ctcaagacct gcaagtggct tctgctcaag caattgggtg agcagtggga      420
cgcaatcctt ggtctatcct agtaccttgt catcgttgtt tgggagcagg caagcgtctg      480
acaggttatg ctgcaggagt ggaaaaagaa gcttggctct tggagcatga aggagtagat      540

```

tttaaagata gaagcaatag aaggagaagc acatgttag 579

<210> 128  
 <211> 1455  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 128  
 aggaattgtg gattgccaaa ttgtatcatt gaaattattg ctcaaatttg ttatgatata 60  
 aatatgaata aaagtagact aggacgtggc agacacggga aaacgagaca tgtattattg 120  
 gctttgattg gtatttttagc aatttctatt tgcctattag gcggatttat tgcttttaag 180  
 atctaccagc aaaaaagttt tgagcaaaag attgaatcgc tcaaaaaaga gaaagatgat 240  
 caattgagtg agggaaatca gaaggagcat tttcgtcagg ggcaagccga agtgattgcc 300  
 tattatcttc tccaagggga gaaagtgatt tectctgtta gggagctgat aaatcaagat 360  
 gtaaggaca agctagaaa taaggacaat cttgttttct actatacaga gcaagaagag 420  
 tcaggttaa agggagtcgt taatcgtaat gtgaccaaac aaatctatga tttagtgtgt 480  
 ttaaagattg aagagactga aaagaccagt ctaggaaaagg ttcacttaac agaagatggg 540  
 caaccttcta cacttgacca actgttttca gatcctagta aggctaagga acagctgata 600  
 aaagagtga cctccttcac agaggataaa aaaatagagc aagaccagag tgagcagatt 660  
 gtaaaaaact tctctgacca agacttgtct gcatggaatt ttgattcaa ggatagtcag 720  
 attatccttt atccaagtcc tgtggttgaa aatttagaag agatagcctt gccagtatct 780  
 gctttctttg atgttatcca atctctgtac ttactcgaaa aagatgcggc cttgtacca 840  
 tcttactttg ataagaaca tcaaaaagtt gtcgctctaa cttttgatga tggtcctaat 900  
 ccagcaacga ccccgaggt attagagacc ctactctaat atgatattaa agcgactttc 960  
 tttgtgcttg ggaataatgt ttctgggaat gaggacttgg tgaagaggat aaaaatctga 1020  
 ggtcatgttg ttggaacca tagctggagc catcgatct tctcgcaact ctctcttgat 1080  
 gaagctaaaa agcagattac tgatactgag gatgtgctaa ctaaagtgtc ggggtctagt 1140  
 tctaaactca tgcgtccacc ttatgggtgt attacagat atattcgcaa tagcttggat 1200  
 ttgagcttca tcatgtggga tgtggatagt ctggactgga agagtaaaaa tgaagcatct 1260  
 attttgacag aaattcagta tcaagtagct aatggctcta tcgttttgat gcatgatatt 1320  
 cacagtcgga cagtcattgc cttgccagg gtcattgagt atttgaaaa tcaaggttat 1380

acctttgtga ccataccaga gatgctcaat actcgccataa aagctcatga gctgtactat 1440  
 agtcgtgatg aataa 1455

<210> 129

<211> 744

<212> DNA

<213> *Streptococcus pneumoniae*

<400> 129

ctacggtttt tatttgtata tggtagaatc tttttacaaa aatacttggg aatcttgttt 60  
 attcatgcta taataggaac aattactttt aggagggtgca gtatgtctta ttbatttgag 120  
 atattaccga gtttactgaa tgggtgcgagc acgactgtac aggtcttttc actggtcttg 180  
 ctatttttcca ttcccttggg cgtttttgatt gcctttgcct tgcaagtcca ttggaagccc 240  
 ctccattatc tgattaacat ttacatctgg gttatgcgag gaacccctt actcttgcaa 300  
 ctgattttta tctattatgt gctcccaagt attgggattc gtttagaccg ccttcctgca 360  
 gctattattg cctttgttct caactatgca gcttactttg cagaaatttt ccgtggggga 420  
 attgacacta ttccaagagg acagtatgag gccgccaagg tcttgaagtt tagccctttt 480  
 gacagagtgc gctatattat cttgccccaa gtgaccaaga tcgttcttcc tagtgtcttt 540  
 aatgaagtta tgagtttggg caaggatact tctttgtct atgctctcgg aatttcagac 600  
 cttatcttgg ctagtccaac agctgctaac cgcgatgcta gtctagttcc tatgtttctg 660  
 gcaggagcca ttbatttgat ttgtattggg attgtgacaa ttattttcaa aaaagttgag 720  
 aagaagtata gttattatag atag 744

<210> 130

<211> 717

<212> DNA

<213> *Streptococcus pneumoniae*

<400> 130

atggaagaaa gtattaatcc aatcatctct attggtcctg ttatcttcaa tctgactatg 60  
 tttagcatga ctttgttgat tgtgggagtt atttttgtct ttattttattg ggcaagccgc 120  
 aatatgacct tgaacccaa aggaagcaa aatgtacttg agtatgtcta tgactttgtt 180  
 attggattta cagaacctaa cattgggttcg cgctacatga aagattactc actctttttc 240  
 ctttgtttat tccttttcat ggtgattgcc aataaccttg gcttaatgac aaagcttcaa 300

```

acgatcgatg ggactaactg gtggagttcg ccaaccgcta atttacagta tgacttaacc 360
ttatcttttc ttgtcatttt gttgacacat atagaagcgc ttgcgcgcgc tggatttaaa 420
aaaagataaa aatcttttat gagtccctgt tttgtcatac cgatgaatat cttggaagaa 480
tttacaact tcttatcttt ggctttgcgc atttttggga atatctttgc aggagaggtc 540
atgacagatt tgttactctt tctttccac caagctattt attggtatcc agtagccttt 600
ggagctaatt tggcttggac tgcattttct gtctttattt cctgcatcca agcttatgtt 660
tttactcttt tgacatctgt gtatttaggy aataagatta atattgaaga ggaatag 717

```

```

<210> 131
<211> 1695
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 131
gataataat tatgattat caacaaggag gaaaaacttt tgagtgaata gtcaagagaa 60
gaagagaat taagcttta agagcagatt ctgagagatt tagaaaaagt aaagcgtat 120
gatgaagttc tgaaagaaga tgaggcagta gttgcgactc ctgcaaatga accttcaact 180
gaagaactca tggctgattc cttgtcaacg gtgaggaga ttatgagaaa agctcctacc 240
gtgcctactc acccaagtca aggtgtacca gcttctccag cagatgagat tcaagagaa 300
actcctggtg ttccaagtoa tccaagtaa gatgtacctt cttctccagc ggaagaaagt 360
ggatcaagac caggtccagg tctgtttaga cctaagaaac ttgaaagaga atacaatgaa 420
acccaacaa gggtagctgt ttctatacg acggcagaga aaaaagcaga acaagcaggt 480
ccagaaacac ctacgcctgc tacagaaaca gtggatatca tcagagatac atcacgtcgt 540
agccgtagag aaggagcaaa acccgtaaag cctaagaaag agaagaagtc acatgtgaaa 600
gcttttttga ttctattcct tgtattcctt gccttgctct cagcaggttg ttactttggt 660
taccagtagc tgctagattc cttattacct atcgatgcta attctaagaa atatgtgacg 720
gttggaattc cagaaggttc aaacgttcaa gaaatcggtc cgacgcttg aaaagctggt 780
ttggtaaagc atggtctgat ttttagtttt tatgccagt ataaaaatta taccgacttg 840
aaagcaggtt actacaattt gcaaaagagt atgagtacag aagacttact caaagagttg 900
caaaaaggtg gaacagatga accgcaagaa cctgtacttg cgactttgac aattccagaa 960
ggttatacct tggatcagat tgcctcaagct gtgggtcaat tgcaaggtga cttcaagag 1020

```



tctttgacag cggaggcttt cttggctaaa gttcaagatg agacgtttat cagtcaagca 1080  
 gtagcgaat atctactttt actgaaagt ttgctgtaa aagacagcgg tgcgcgttat 1140  
 cgtttggaag gatacctttt cccagctaca tactctatca aggaagcac aactattgag 1200  
 agcttgattg atgagatgtt agctgctatg gataagaacc tatctcctta ctatagtact 1260  
 atcaaactta aaaacttgac tgtcaatgag ttgttgacca ttgcttcctt ggtcgaaaaa 1320  
 gaaggtgccca agacagaaga tcgtaagctc attgcagggtg tattctacaa togtttgaat 1380  
 cgtgatatgc cacttcaaag taatattgca atcttgtatg cccaaggaaa actggggcaa 1440  
 aatatcagtc tagctgagga tgttgcgatt gataccaaca ttgattcacc ttataatgtt 1500  
 tataaaaatg tagtctcat gcctgggtcca gtcgatagtc caagtctgga tgcgattgag 1560  
 tcaagcatca atcaaaactaa gagcgataac ctctactttg tagcagatgt cacagaaggc 1620  
 aagggtctact atgctaacaa tcaagaagac cagaccgca atgtcgctga acatgtcaac 1680  
 agcaaattaa actaa 1695

<210> 132

<211> 879

<212> DNA

<213> Streptococcus pneumoniae

<400> 132

tcgtcgttta caggaggaaa tttaacaggg caattgactg aaaagattca agaacatgaa 60  
 ttaattaaga ctaaccaagc agagaaaagt gtacaggatg ttttgataa ttgtattgaa 120  
 aggttacaac acaattcact gaaatcagat agggttactt cttttgagac cccgtttgct 180  
 ctcttattta tctttcgcac tatagctgtg atgctaacct atgggggtta tcgggtcagc 240  
 gcaggatata tatctgtggg aaccttgggt tcgtttttga ttacctctt tcaattactt 300  
 aatcctatta gtaatatagc taattttgta actgtttatt ctaggagcaa gggatcttca 360  
 gttgcactgg agaacttgct tgcagttcct aaagaaaaat ttgagggagg aaaaatcgta 420  
 tcaggacgag ggttgaaatt taacctgtgc tatttttggt atgatgaaaa tcgacctgtc 480  
 tttaaaggata ttacttggtc aattttcaag gggcaaaaaa ttgcttttgt tggacctct 540  
 ggatcaggaa aatcaacgat tgtgcgtttg tttagcggtt ttataaacc gctttcagga 600  
 gatattctaa tggagcaatc aagtatatat gattttaact taaaagaatg gagaagtaaa 660  
 atcgcttggg ttccacaaaa taatgcagtc ttatctggca gtattcgtga caatctttgt 720

ctcgggttga atcgcttagt aactgatgat gaattgatga aagtgc taga cttagtatca	780
ctagggtgatg agattcgcct catgaaagag ggactagata ctgaagttgg tgaacgcgga	840
cgactcttgt caggggggagc aacgaaagac ttcaaatag	879

<210> 133  
 <211> 555  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 133	
ggagtattta tgaattataa attattaaga gtagatacta aggtgattat ggggagtttc	60
ttacttggtc tgtctagtct acttgctttg ttgcttcccc ttactctaaa ggatttaata	120
gatgggaggt ctattgaaaa tataggctcc aaagtatttc aatcgttttt gattttttatt	180
ggccaagcct tgttttcttc tattgggtac tatctgttta gtcaatcggg tgaaaaaaag	240
atagcaaaaa tcaggaaaaa agtgatagag ggggtgattt atgtagagaa atccttcttt	300
gataagagcc aaagtgggga gttgacttct gccattgtca atgacacgag tgtcattcgt	360
gagtttttaa ttacgacttt cccaaatatt attctgagtt tagttatggt acttggttcc	420
attgtagtct tatttagtct tgattggaat ctttctctac ttttattcat cactcttctt	480
tgtatgatgt ttattatctt gccctttcc aatatcagtg aaaagtatag tcgtcgttta	540
caggaggaaa tttaa	555

<210> 134  
 <211> 989  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 134	
aacaaatatt taaagcagga ggttcggaa atgaaaaagt ctaagagcaa atatctaacc	60
ttggcaggtc ttgtcctggg tacaggagtt ttattgagcg cgtgtggaaa ttctagcagc	120
gcgtcaaaaa cctacaacta tgtttattca agtgatccat ctactgttaa ctatctagca	180
gaaacccgcg cagcaacatc cgatattgtt gcaaatttgg tagacggggt attagaaaat	240
gaccaatatg ggaatattat tccatcatta gcaggagatt ggactgtttc tcaggacggt	300
ttgacctata cctacaactc tcgtaaggat gccaaagtgt ttacttttga gggagaagaa	360
tatgcgcctg taactgccca ggattttgtg acaggtttgc aatatgcagc tgataaaaaa	420
tcagaagcct tgtatctagt gcaggactct gttgctggtt tggatgacta tatcactggt	480

```

aaaacaagcg acttttcaac tgtcgggtgtc aaggcacttg atgaccaaac ggttcaatat 540
actttgggtta aaccagaact ttactggaat tcaaaaacac ttgcaacgat actttttcct 600
gttaatgcag atttctgaa atcaaaaggg gatgattttg ggaaggcgga tccatctagt 660
attttgtaca atggaccttt cttgatgaaa gcacttgtct caaatctgc tattgaatat 720
aagaaaaacc ctaattactg ggaatgctaa aatgtctttg tagacgatgt gaaattgacc 780
tactatgatg gtacgcacca agaactcactg gaacgtaatt ttacagctgg tgccttact 840
acggctcgtc tttttcctaa cagctccagc tatgaaggga ttaaagaaaa atacaaaac 900
aatatcatct atagtatgca aaattcaact tcataatttct ttaattttaa cctagatagg 960
aagtcttaca attatacttc taaaacaagt gacattgaaa agaatacgac tcaggaagca 1020
gttctcaata aaaacttcog tcaggctatc aattttgctt ttgacagaac atcttatggg 1080
gctcagctcg aagggaaga aggtgcaaca aagattttgc gtaacctagt ggttctcca 1140
aactttgtca gtatcaaggg aaaagacttt ggtgaagttg tagcctctaa gatggtaac 1200
tatggtaagg aatggcaagg tatcaacttt gcggatggtc aagaccctta ctacaacct 1260
gagaaagcca aggctaagtt tgcggaagct aagaaagaac tcgaagcaaa ggggtgtcaa 1320
ttcccaatcc acttgataa gactgtggaa gtaacagata aagtaggcat acaaggagt 1380
agttctatca aacaatcaat tgaatctgtt ttaggttctg ataatgtagt gattgacatt 1440
cagcaattaa catcagatga gtttgacagt tcaggctact ttgctcaaac agctgctcag 1500
aaagattatg atttatatca tggcggttgg ggacctgatt atcaagacct gtcaacctat 1560
ctcgatatth ttaatactaa tagtggagga tttctgcaaa atcttggaact agagcctggt 1620
gaggccaatg acaaggctaa ggcagttgga ctggatgtct atactcaaat gttggaagaa 1680
gctaataaag agcaagatcc ggcacaacgt tatgagaaat atgctgatat tcaagcttgg 1740
ttgattgata gttctttagt tcttccaagt gtttcgcgtg ggggaacacc atcattgaga 1800
agaaccgtac cattttgtgc tgacctatgg ttaaccggtc caaaaggggt tgaatcatat 1860
aaatacctca aagtacaaga taagattgtc acaacagacy aatatgcaaa agccagagaa 1920
aatgggttga aagaaaaaga agaattcaat aaaaaagccc aagaagaatt ggcaaacat 1980
gtcaataaa 1989

```

&lt;210&gt; 135

<211> 1647  
 <212> DNA  
 <213> *Streptococcus pneumoniae*

<400> 135  
 ttatcaaaat tgaatgagga atctatgtcg cacgaaaaca atcaccagca gcccagatg 60  
 ttacggggga ctgcttggtc aacggctagt aactttatca gtgcctact cggggctggt 120  
 tacattatcc ctgggtacat ctggatgggg gcttatgcag ctaaggcaaa tgggtctctt 180  
 accatgggtt acaatatcta tgcttggttc ttgttggtt caacagcggg gattocagtt 240  
 gcgggtggcca agcaagttgc caagtataat accatgcgag aagaagagca tagctttgcc 300  
 ctgattcgga gcttcttagg ctttatgaca ggactaggcc tgggtttttgc tttagtcttg 360  
 tatgtctttg ctcttggtc agcagacttg tctggcgtgg gcaaagactt gatcccaatc 420  
 atgcaaaagt tggcttggtg agtcttgatt ttcccgctta tgagtgttat ccgaggattt 480  
 ttccaaggga tgaataacct caaacctat gccatgagcc aaattgctga gcaggctatt 540  
 cgtgttatct ggatgtctct agcaaccttt atcattatga agctcgggtc aggagattat 600  
 ctagcagccg ttaccaatc aacctttgct gcctttgtcg gtatggtagc cagttttgca 660  
 gtcttgattt atttcttgc ccaagaaggc tcaactcaaa gaatctttga aacaggagat 720  
 aagattaaca gtaagcgtct ctgggttgat accattaagg aagccattcc ttttatcctg 780  
 acagggtctg ccattccagct ctccagatt ttggatcagc tgacctttat caatagtatg 840  
 agctgggtta ccaactacag caatgaggac ttggttgtca tgtttctta ttctcagcc 900  
 aatcctaata aaatcacgat gattttgatt tctgtagggg ttccgattgg gagtgttggt 960  
 ttgccacttt tgacggaaaa ctatgtcaag ggggacttga aagcagcttc tcgtctcggt 1020  
 caggacagtc tcacctact ctttatgttc ttgtaccag caacggttgg agtggttatg 1080  
 gtaggagaac ctctttatc ggtcttctat ggtaagccag atagtttggc tctgggtta 1140  
 tttgtctttg cagttttgca gtctattatt tttaggttgt acatgggtct gtctccaatg 1200  
 cttcaggcca tgttcgcaa ccgcaaggcc gtctctatt ttatctatgg ttctattgcc 1260  
 aagctagtct tgcaactacc taccatgcc ctcttcaca gttatgggcc tttgatttca 1320  
 acaacattg ctctcatcat tctaacgct ttgatgtatc gggatatttg taaagtaact 1380  
 ggtgtcaagc gcaaggtgat ttgaagca accattttaa tcagtttgc gacctagtc 1440  
 atgtttctgt taataggaac catccagtggt ctgttaggat tttcttcca accaagtgga 1500

cgtttgtgga gcttctttta tgtagctctt gtcggtgcc a tgggggtgg accttatatg 1560  
 gttatgagtc tgcgtacctt tttattagat aaggtaatag gaaaagccca agcagatcgc 1620  
 ctgcgagcaa aatttaagct ttcgtaa 1647

<210> 136  
 <211> 639  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 136  
 gcaaatttct tgcaagttct tttgttttgt tgtaatatat ttataacaa cgagagagtt 60  
 ctgaaattt taagaaaaag gagacacatc atgtctaaaa aagtattatt tatcgtagga 120  
 tcactacgtc aaggttcttt caaccaccaa atggcgctcg aagctgagaa agcacttgct 180  
 ggtaaagcgg aagttagcta ccttgattat tcagcccttc ctctcttcag ccaagatttg 240  
 gaagtccaa cacatccagc tgtagctgct gctcgtgaag cagttctcgt tcgggatgct 300  
 atctggattt tctctccagt ctacaacttc tctatccctg gtacagtga aaacttgctt 360  
 gactggctat ctctgcccct tgacttgctt gatacacgtg gcgtttctgc ccttcaagac 420  
 aagtttgtca cagtatcatc tgtagccaat gcagggcacg atcaactttt cgctatctac 480  
 aaagacctct tgccatttat ccgtacacaa ggcgttgggt atttactgct tgcacgtggt 540  
 aatgactctg cctgggcaag cggaaaattg gttcttgaag aaacagtcct aaactcactt 600  
 gaaaaacaag ctcaagactt ggtcgaagct atcaagtaa 639

<210> 137  
 <211> 1902  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 137  
 agccatccat gcttacctga gggagaaaa atgagtgatt ttatcgttga aaaactaagt 60  
 aaatccgttg gtgacaagac cgttttttag gatatttctt ttattatcca tgacttagac 120  
 agaattggtt taatcgggtg caatgggact ggcaagacca cccttttgga cgctcctttt 180  
 ggtgtttctg gatttgatgg ggaatgcagt cctttttcag ctaaaaatga ttaccagatt 240  
 ggttacttga ctccagatcc tgattttgat gatagaaaga cagttttgga tacgggtcta 300  
 tctagtgaac tcaaggaaat ccagctcatt cgtgagtatg aattgattat gctcgactat 360  
 agtgaggaca agcaggcgcg tttggaacgt gtcatggcag agatgggactc tctccaagct 420

tgggaaatcg aaagtcaggt caagaccgtt cttagcaaat tgggcattca agacttatct	480
actcctgttg ggggaattgtc aggtggtctg agaagacggg tacagttggc acaagtctta	540
cttggcaacc acgacctctt gcttttggat gagccgacca accatctgga tattgcgatt	600
attgagtggtc tgacctcttt tttgaaaaat tctaagaaga cgttcctttt tatcactcac	660
gatcgttatt tcttagacgc tttgtcaaca cggattttcg agttggatcg tgcaggcttg	720
accgagtacc agggaaatta ccaggactat gttcgcttaa aggcggaaca ggatgagcgc	780
gacgcggctc ttcttcacaa aaaagaacaa ctctacaac aagaattggc ctggatgcgc	840
agacaaccgc aggcgcgtgc gaccaagcaa caagctcgta tcaatcgttt ccatgatctg	900
aaaaaggaag tttcaggcag tagtgctgag acagacttga ctatgaactt tgaaccaggt	960
cggattggga agaaagtcatt cgagtttcag gatgtttcct ttgcctatga aaataagccc	1020
attttgcata attttaattct cttagttcag gctaaagacc gtattggaat tgttggggac	1080
aatggtgttg gaaaatcaac cctacttaac ctgattgcag gaagtcttga gccgacagca	1140
ggacaagtgt tgattgggga aactgttcgc atcgctattt tctctcaaca aattgagggt	1200
ttggatgaaa gcaagcgtgt gatcaattac ctgcagggaag tggcagagga ggtaagacc	1260
agtgtgtgtt ctacgacttc catcgctgag ttgctggagc aattcctctt cccacgttcg	1320
acgcattgga ctttgattga gaaattgtca gggggtgaga aaaaacgtct ttaattcctc	1380
aaactgcttt tggaaaaaac aaatgttctt ctttttagacg agccaaccaa tgacctagat	1440
attgcaactt tgacagtctt agagaatttc ttgcaagggt ttgcagggtc cgttttaaca	1500
gtcagtcacg accgctattt cttggataag gtagcgacca agattctcgc ttttgaggat	1560
ggcaagattc gtcctttctt ttgtcattac accgactatc ttgatgaaaa agcttttgaa	1620
acagatatgg ccaattcaagt gcaaaaggcc gaaaaggaaa aagtggtcaa gggttcagaa	1680
gacaagaaac gcatgaccta ccaagaaaag caggagtggg caagtattga aggtgatatt	1740
gaaaccttgg aaaaacgtat cgctgctatt gaagaggaaa tgcagggtcaa cggtctgtac	1800
tttggtgaag ttggtactct ccaaaaaagaa ttggatgaga aaaatgaagc actccttgaa	1860
aaatacgaac gctatgagta tctcagttaa tttgatagtt aa	1902

<210> 138  
 <211> 579  
 <212> DNA

<213> *Streptococcus pneumoniae*

<400> 138

tatactaagg tagtaatcat taagaagtgg ttacaaaaa taatgaatga ggtaaagaaa	60
atggtagaat tgaaaaaaga agcagtaaaa gacgtaacat cattgacaaa agcagcgcca	120
gtagcattgg caaaaacaaa ggaagtcttg aaccaagctg ttgctgattt gtagttagct	180
cacgttgctt tgcaccaagt gcactgggtat atgcatggtc gtggtttctt tgtatggcat	240
ccaaaatagg atgagtacat ggaagctctt gacgggtcaat tggatgaaat cagtgaacgc	300
ttgattacac tcggtggaag cccattctct acattgacag agttccttca aaatagtga	360
atcgaagaag aagctgggtga ataccgtaat gttgaagaaa gcttggaaacg tgtctctgtt	420
atctaccgtt acttgtcaga acttttccaa aaaggtttgg atgtcactga tgaagaaggt	480
gacgatgtga caaacgggtat ctttgcaggc gctaaaactg aaacagataa aacaatttgg	540
atgcttgcag cogaacttgg acaagcacct ggtttgtaa	579

<210> 139

<211> 1083

<212> DNA

<213> *Streptococcus pneumoniae*

<400> 139

tctagcaatc ttttgtttgg gcttatcggc tgcatttatg gggcgtttgg tagaaaaatt	60
tggtccgaaa gtcattggga gtctatctgc ttttctatac gcagggtggaa atatcttaac	120
aggatttgca atagaccgtc agagctgtgg ttgtgtatc tcgcttatgg catttttaggt	180
gggcttggtt tgggagcagg ctatattacc cctgtgtcga cgattataaa atggtttctt	240
gataaacgtg gtctcgaac aggttttagcg attatgggtt ttggttttgc tctcttattg	300
actagtccca tagcgcaaca cctcatcgca ggggtagggc ttgtagaac tttttatatt	360
ttaggagcaa gttactttat tatcatgctc ctactgtcac aattcattaa gcgtccaaat	420
gagcaagagc ttgcaatttt atctcttcca gggaaagaaa aaacagcctc tttagcgcaa	480
ggaatggctg caaatcaggc tctaaaaagc aatcggtttt atatgctttg gattattttc	540
tttatcaaca tagcttgtgg tttaggctta atttcagcgg catcgccaat ggcacaggag	600
atggctggct tgtctacaag tcatgcagca gtaatgggtg gtgttttggg gattttcaat	660
ggatttggct gcttgcctcg ggcaggtttg tctgactata tcggctcgcc tctaaccttt	720
agtatattac tgctgtgcaa tcttttcttt tctctctcac ttggctctt tacagattcc	780

```

gttttatttg tagttgctat gtctattttg atgacttgct atggagctgg ttttcttttg      840
attccagctt atctcagtgga tattttttgga accaaggaat tggcgcgtct gcattggttat      900
attttaacag cttgggcaat ggctgggtta gggggaccta ttttattagc agagacttat      960
aaaatggctc attcgtacac acaaaccttg ttctgttttc tcattttata cagtatcgcc      1020
ttggcctttgt cttattatct aggtcgttca atcaaaaaag aaagtcaaaa agcgcttaca      1080
tga                                                                           1083

```

```

<210> 140
<211> 468
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 140
gacaatatga agcaaacaaa aacaactaaa atcgcccttg tatccctatt aacgcacctt      60
tctgtggttc taggttattt cttaaaaaat ccaacacctta caggaattct aactctttta      120
gatgctggtg tcttctttgc ggccttttac tttggtagtc gtgaaggagc ggtagtcgga      180
ggactagcaa gtttcttgat tgacctctta tcaggctacc ctcaagtggat gttcttttagc      240
ttggtaacc atggcttgca gggatttttc gcaggattta aaggaaaaag tcagtgggta      300
ggcctatttt tagcaactat tgccatggta ggaggctacg ccttgggttc tgctttgatg      360
aatggctggg cagcagccct ccagaaaatt ctaccgaatt ttatgcaaaa tatgtagtagg      420
atgattgtag gatttattct tagtcaaagt atcaagaaga ttaagtaa                    468

```

```

<210> 141
<211> 684
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 141
gagaagatga tttcaagag attagaattg gtagcttcct ttgtgtcaca gggggctatt      60
ttactagatg tgggaagtga ccatgcttat ctgcctatcg agttgggtga gagaggccaa      120
atcaaaacgc ctattgcagg tgaggtggtg gaaggtccct atcagtctgc ggttaaaaat      180
gttgaggctc acggcctaaa ggagaaaatc caagtccgtt tagccaatgg cttggcagct      240
tttgaagaga ctgaccaagt gtctgtcatt accattgctg gcattgggtg tcgtttgatt      300
gctaggattt tagaagaagg tttggggaag ttagctaatg tagagcgttt gatcctccag      360

```



cccaataatc gtgaagacga cttgcgtatc tggctacagg atcatggatt ccagattgta 420  
 gcagaaagca tcttagaaga agctggaaag ttttatgaga ttttggtggt ggaagcagga 480  
 caaatgaagc tatcagccag tgatgttcgc tttggtccct tcttgccaa agaagtcagt 540  
 ccagtatattg tccaaaaatg gcaaaaagaa gctgagaagc tagagtccgc cctcgggaca 600  
 atcccagaaa aaaatctgga agaacgtcaa gttctagtag ataagattca agctatcaag 660  
 gaggtgctcc atgttagcaa gtga 684

<210> 142  
 <211> 336  
 <212> DNA  
 <213> *Streptococcus pneumoniae*

<400> 142  
 gaaaaaattt tggagggtat ccgtatgaaa attggtggtg ttgcagcttg tactgtggga 60  
 attgccaca cttatattgc acaggaaaaa ttgagaatg ccgcaaaggc agctggacat 120  
 gtgattcatg ttgagactca ggggacaata ggggtagaaa atgaattgag tcaagagcag 180  
 attgatgcag cggatgtagt tatttttagca gttgatgtta agattttctg tatggaacgc 240  
 tttgagggta aaaagattat caaggttcca acagaagtgg cagtcaaacc tcccaataaa 300  
 ctgattgcta aagctgttga gattgttacg aaataa 336

<210> 143  
 <211> 777  
 <212> DNA  
 <213> *Streptococcus pneumoniae*

<400> 143  
 cttgacttaa tttttttttt aatgtatatt aagagacagg aggaatacaa gtttatgata 60  
 cgtatcgaaa acctcagtgt ctctacaaa gaaacgttgg cacttaagga tatttacta 120  
 gtgctccatg gaccaacaat taccggcato attggtccaa acggcgtctg gaaatcaaca 180  
 ctattaaaag gtatgttggg aattatocca catcaaggc aggcatttct cgtagacaag 240  
 gaagttaaaa aatccttaca ccgaattgcc tatgtcgaa aaaaatcaa tatcgactac 300  
 aactttccca tcaaggtcaa ggaatgcgtc tcgttaggac tatttccttc tattctcttc 360  
 tttogaagtt taaaggctaa acattggaag aaagtgcagg aggccttga aatcgctggc 420  
 ctagctgact acgctgaacg tcaaattagt caactgtctg gaggtcaatt ccagcgggtc 480  
 ttgattgcca gatgtttggt gcaggaagcc gactatatcc tcttgatga acccttggct 540

gggattgact ctgtcagtga ggaatcatc atgaatacgc tgagagattt gaaaaaagct 600  
 gggagacgg ttctcatcgt tcaccacgac ctacagcaaga ttcccacta cttcgcataa 660  
 gtcttacttg tcaatcgaga agtgattgcc ttgtgtccaa caaaagaac ttttaccgaa 720  
 accaatctaa aagaagctta cggtaatcaa ctctttttca atggagggtga cctatga 777

<210> 144  
 <211> 897  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 144  
 aaggaggat ttatgacata ttacgttgca attgatatog gtggaacca catcaagtat 60  
 ggtttggttg atcaagaggg gcaactctct gaatcgcatg aaatgccaac tgaggcgcat 120  
 aaggggtggac ctcatatctt acaaagacc aaagatatcg tagctagtta tttgaaaaa 180  
 ggcccagtag caggtgttgc catatcttct gctgggatgg tggatccgga taagggtgag 240  
 attttctatg tctggccgca aatccctaac tacgcaggca ccaggttcaa aaaggaatc 300  
 gaagaaagct ttactattoc ttgtgagatt gaaaatgatg tcaactgtgc aggtcttgct 360  
 gaggcagtat ctggttcagg caaggagaca agtggtgacac ttgtctgac cattggaacc 420  
 ggtatcgggtg gttgcttgat tatggatagg aaagtcttcc atgggttttag caattcagcc 480  
 tgtgaagtcg ggtatatgca tatgcaggat ggagcttttc aagacttggc ttctacaaca 540  
 gctttagtga aatatgtagc tgaagcccat ggagaagatg ttgatcagtg gaatggccgt 600  
 agaattttca aagaagccac tgaaggaaac aaaatctgca tggagggtat tgaccgtatg 660  
 gttgactatc taggaaaagg tctggcaaat atttgctacg ttgccaatcc agaagtggtt 720  
 attcttggtg gtggtatcat ggggcaagag gctatcctca aacctaatag ccgtacagcc 780  
 ttgaaagagg ctttggtagc aagtttagca gaaaaaacac gattagaatt tgcccatcac 840  
 caaaatacag cagggatggt ggggtgcatat tatcatttta agacaaaaca atcctag 897

<210> 145  
 <211> 690  
 <212> DNA  
 <213> Streptococcus pneumoniae

<400> 145  
 caaaaaagaa aacagtttac aaagaaaaat gatggaggag caaacatggc acaaaaagga 60

gtaagcctta tcaaggcagc atttgataca gataactttc tcatgcgttt tagtgagaag	120
gtcttggaca tctgtacacg caatcttctt ttgtctgtct cttgtttacc catcgtgacg	180
attggagtgg ctaaaatcag cctctacgag accatgttcg aagttaagaa gacgacagcg	240
gtgcctgttt ttaaaatcta tctaagatct ttcaagcaaa atctgaaact aggtcttcag	300
ctgggtttaa tggagtttag aattgtgttt cttacccttt cagatctcta tctttctcgg	360
ggtaaacag ctctgcctt ccaattgctg aaagccattt gtttaggtat totgattttt	420
cttactatcg tgatgctggc tagttacct atcgcggcac gttatgacct atcttggaaa	480
gaaattcttc aaaaaggatt gatgttggct agttttaact ttccttgggt cttctctcag	540
ttagccattc ttgtctcat tgtgatggtt ctttatctgt cgccttcag tctactctta	600
ggtggctcag tcttctact ttttgggtt ggactattgg tctttatcca gactggattg	660
atggagaaaa ttttcgcaaa ataccaatag	690

&lt;210&gt; 146

&lt;211&gt; 915

&lt;212&gt; DNA

<213> *Streptococcus pneumoniae*

&lt;400&gt; 146

cccaatcttg taaaagaagg gagaaggaga atggttaaag aacgtaattt aactcgtcgg	60
atatttgttt tgccagctat gattatcgta ggattactct ttgtttatcc gttttctcgg	120
agtatttttt atagctttac caataagcat ttgattatgc ctaattataa atttgttgggt	180
ttggctaact ataaagctgt gctatcagat cccaacttct ttaatgcgtt ctttaattca	240
attaagtgga cgtttttctc attagtgtgt caagtttttag tagggtttgt attggcttta	300
gctcttcaca gagtacgcca cttcaagaaa ttatatagga cattattgat tgttctctgg	360
gcatttccta ccatcgttat tgccttctct tggcagtgga ttctaaacgg ggtttatggc	420
tacttacota atctaactgt aaaattagggt ttaatggaac atacacctgc atttttgaca	480
gatagtacat gggcattcct atgtttgggt tttatcaaca ttgggtttgg agccaacatg	540
attatgggta atgtgcttct agctttgcaa acagtaccag aagaacaatt tgaggctgct	600
aagatagatg gtgcttcaag ttggcagggt ttcaagttaa tgcgtcttcc acatattaaa	660
gtggtttagt gacttctagt tgttttgaga actgtatgga tctttaataa ctttgacatt	720
atctacctca ttaactgggtg tggaccagcc aatgctacaa cgacgcttcc aatttttgct	780

tacaacctgg gctggggaac taaattgttg ggtcgtgctt cagcagttac agtactgctc	840
tttatctctct tgggtggcgat ttgctttatc tactttgcta tcatcagtaa gtgggaaaag	900
gagggtagaa aataa	915
<210> 147	
<211> 1356	
<212> DNA	
<213> Streptococcus pneumoniae	
<400> 147	
tgtagaaaga gaagaacgat gaaaaaaatg agaaagtttt tatgtctagc tggaaatgcg	60
ctagcggtctg ttgccttggt agcttggtca ggaaaaaag aagctacaac tagtactgaa	120
ccaccaacag aattatcttg tgagattaca atgtggcact cctttactca aggacccggt	180
ttagaaagta ttcaaaaatc agcagatgct ttcatgcaaa agcatccaaa aacgaaaatc	240
aagattgaaa cattttcttg gaatgacttc tatactaaat ggactacagg tttagcaaat	300
ggaaatgtgc cagatatcag tacagctctt cctaaccaag taatggaaat ggtcaactca	360
gatgcttttg ttccgctaaa tgattctatc aagcgtattg gacaagataa atttaacgaa	420
actgccttaa atgaagcaaa aatcggagat gattactact ctgttctctt ttattcacat	480
gcacaagtca tgtgggttag aacagatttg ttaaaagaac ataattattga ggttctctaa	540
acttgggac aactctatga agcttctaaa aaattgaaag aagctggagt ttatggcttg	600
tctgttcggt ttggaacaaa tgacttaatg gcaacacggt tcttgaactt ctacgtacgt	660
agtgtgggag gaagcctctt aacaaaagat cttaaagcag acttgacaag ccaacttgct	720
caagatggta ttaaatactg ggttaaattg tataaagaaa tctcacctca agattctttg	780
aactttaatg tcttcaaca agctaccttg ttctatcaag gaaaaacagc atttgacttt	840
aactctggct tccatctcgg aggaattaat gccaacagtc ctcaattgat tgattcgatt	900
gatgcttctc ctattccaaa aatcaaagag tctgataaag accaaggaaat tgaaacctca	960
aaacattccaa tgggtgtttg gaaaaattca aaacatccag aagttgctaa agcatcttta	1020
gaagcacttt ataattgaaga agactacggt aaattccttg attcaactcc agtaggtatg	1080
ttccaacta ttaaggggat tagcgattct gcagcctata aagaaaatga aactcgttaag	1140
aaatttaaac atgctgaaga agtaattact gaagctgta aaaaagggtac tgctattggt	1200
tatgaaaatg ggccaaggtg acaagctggt atgttgacta accaacacat tattgaacaa	1260

atgttccaag atatcattac aaatggaaca gatcctatga aagcagcaaa agaagcagaa 1320  
 aaacaattaa atgatttatt tgaggctggt cagtag 1356  
  
 <210> 148  
 <211> 2403  
 <212> DNA  
 <213> *Streptococcus pneumoniae*  
  
 <400> 148  
 atgtcttatt tcagaaatcg ggatatagat atagagagga tcagtatgaa tcggagtgtt 60  
 caagaacgta agtgctgtta tagcattagg aaactatcgg taggagcggg ttctatgatt 120  
 gtaggagcag tgggtatttg aacgtctcct gttttagctc aagaaggggc aagttagcaa 180  
 cctctggcaa atgaaactca actttcgggg gagagctcaa ccctaactga tacagaaaag 240  
 agccagcctt ctccagagac tgaactttct ggcaataagc aagaacaaga aaggaagat 300  
 aagcaagaag aaaaaattcc aagagattac tatgcacgag atttggaana tgctgaaaca 360  
 gtgatagaaa aagaagatgt tgaaccaat gcttcaaatg gtcagagagt tgatttatca 420  
 agtgaactag ataaactaaa gaaacttgaa aacgcaacag ttcacatgga gtttaagcca 480  
 gatgccaaag ccccgacatt ctataatctc ttttctgtgt caagtgtac taaaaagat 540  
 gagtacttca ctatggcagt ttacaataat actgctactc tagaggggcg tgggtcggat 600  
 gggaaacagt tttaacataa ttacaacgat gcacccttaa aagttaaacc aggtcagtg 660  
 aattctgtga ctttcacagt tgaaaaaccg acagcagaac tacctaaagg ccgagtgcgc 720  
 ctctacgtaa acgggggtatt atctcgaaca agtctgagat ctggcaattt cattaaagat 780  
 atgccagatg taacgcagt gcaaatcgga gcaaccaagc gtgccaacaa tacggtttgg 840  
 ggggtcaaat tacagattcg gaatctcact gtgtataatc gtgttttaac accgaagag 900  
 gtacaaaaac gtagtcaact ttttaaacgc tcagatttag aaaaaaact acctgaagga 960  
 gcgggtttta cagagaaaac ggacatattc gaaagcgggc gtaacggtaa cccaataaa 1020  
 gatggaatca agagtattcg tattccagca cttctcaaga cagataaagg aactttgac 1080  
 gcagggtcag atgaacgcg tctccattcg agtgactggg gtgatatcgg tatggtcac 1140  
 agacgtagtg aagataatgg taaaacttgg ggtgaccgag taaccattac caacttacgt 1200  
 gacaatccaa aagcttctga cccatcgatc gggtcaccag tgaatatcga tatgggtgtg 1260  
 gttcaagatc ctgaaaccaa acgaatcttt tctatctatg acatgttccc agaaggggaag 1320

```

ggaatctttg gaatgtcttc aaaaaagaa gaagcctaca aaaaatcga tggaaaaacc 1380
tatcaaatcc tctaccgtga aggagaaaaa ggagcttata ccattcgaga aaatgggtact 1440
gtctatacac cagatgggtaa ggcgacagac tatcgcgttg ttgtagatcc tgttaaacca 1500
gcctatagcg acaaggggtga tctatacaag ggtgaccaat tactaggaaa tatctacttc 1560
acaacaaaca aaactttctc atttagaatt gccaggata gctatctatg gatgtcctac 1620
agtgatgacg acgggaagac atggtcagct cctcaagata ttactccgat ggctaaagcc 1680
gattggatga aattcttggg tgtaggctct ggaacaggaa ttgtacttcg gaatgggcct 1740
cacaggggac ggattttgat accggtttat acgactaata atgtatctca cttagatggc 1800
tcgcaatctt ctctgtctat ctattcagat gatcatggaa aaacttgcca tgcctggaga 1860
gcggtcaacg ataaccgtca ggtagacggt caaagatcc actctctac gatgaacaat 1920
agacgtgcgc aaaatacaga atcaacggtg gtacaactaa acaatggaga tgttaaacct 1980
tttatgcgtg gtttgactgg agatcttcag gttgctacaa gtaaagacgg aggagtgtact 2040
tggtgagaag atatacaacg ttatccacag gttaaagatg tctatgttca aatgtctgct 2100
atccatacga tgcacgaagg aaaagaatac atcatcctca gtaatgcagg tggaccgaaa 2160
cgtgaaaatg ggatgggtcca cttggcagct gtcgaagaaa atgggtgagtt gacttggctc 2220
aaacacaatc caattcaaaa aggagagttt gcctataatt cgtccaaga attaggaaat 2280
ggggagtatg gcatcttgta tgaacatact gaaaaggac aaaatgccta taccctatca 2340
tttagaaaat ttaattggga atttttgagc aaaaatctga tttctcctac cgaagcgaac 2400
tag 2403

```

```

<210> 149
<211> 636
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 149
acgatgagac ttgaaattat aaatggacag aaaatttatg ggaaaagacc tatttttaa 60
cagttgaatt tgggtgtttca atcaggaaaa atttatggac ttaaagggtga taatggatct 120
ggcaagacgg ttctttttaa gatacttgct gggttatatta agcttgacaa aggaaaagt 180
cttcagatg gtaaagttta cggggtaaaa aatcattata ttcaggatgc aggaatttta 240
attgaaaag tcgagttttt atctcattta tccctgagag aaaatttggg actgttaaag 300

```

```

tatttttcat ctaaagttac ggaaaaaaga attgcctatt ggattcaata ctatgattta 360
caggaatttg aagacattga ataccgtcat ttatccttag gaacaaagca aaaaatggcc 420
ttgattcaag cctttatttc ctctccttct atactcttct tcgatgaacc tatgaatgct 480
ttggatgaga agagtgtgag gttaaccaa caggtcattt tatcttacct gaaaaagaa 540
aatggtctgg ttatcctgac gtgcacata tcggaagata tttcagacct ttgtacagat 600
gtattagttg tcgaaaatgg acatatata atgtaa 636

```

```

<210> 150
<211> 297
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 150
cggatgcggt ccatgacccg tctggcttcc caggtttcgt catttccatg tttcactttc 60
gcaaaatgct tctccaaatc ttcaaagtgt aagttggatg tgaaaaaggt cggtaaattt 120
tctgcctacc gatattggag aatgacctgc aggtattcgt cagcaccaca aacggttgat 180
tgctcgccgc caatatcctc taaaatcagg acctcagaca gcttaatctc atccaccaag 240
gtcttaacat tgccatcact gatagcattt ttgacatcaa tgacaaagct aggatag 297

```

```

<210> 151
<211> 1509
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 151
tcagtgatta tattaagga gtttaagcct atgtcattac tagtatttga aaatgtatcc 60
aatcatatg gagcaacacc agcccttgaa aatgtttctc ttgacattcc agctggaaaa 120
attgtcgccc tctttgggcc aaacggctca ggaaaaacaa ccttgattaa actaettaat 180
ggcctcttac aaccagatca aggacgtgtc ctcataacag acatggaccc aagcccagca 240
accaaggccg ttgtagctta ttgcctgat acgacctatc tcaatgagca aatgaaggtc 300
aaagaagccc taacctactt caagaccttc tataaagatt tcaatcttga acgcgcccat 360
catctacttg cagacctggg cattgatgaa aatagtcgtc tcaagaaact atcaaaagaa 420
aacaagaaa aggttcaact gatttttggt atgagcgtg atgctcgtct ctatgttttg 480
gacgaaccca ttggtggggt ggatccagca gcccggtgct atatcctcaa taccattatc 540
aacaactact caccaacttc taccgttttg atttctacce acttgatttc tgatatcgag 600

```

```

ccaatcttgg atgaaattgt ctccctaaaa gacggaaaag tcgtccgtca aggaaatgta 660
gatgatattc gctacgagtc aggtgaatcc attgaccaac tcttcogtca gaatttaagg 720
cctaagcaaa ggagattatt tatgttttgg aatttagttc gctacgaatt taaaaatggt 780
aacaagtggg atttagccct ctacgcagcc gtgctagtc tttctgcct catcggaata 840
cagacacaag gctttaaaaa tctaccttac caagaaagtc aggtactat gctacttttt 900
ctagctacag tctttgggtg cttgatgctt acacttggga tttaacccat tttcttgatt 960
attaacgct tcaaaaggtag tgtctacgac cgacaaggct atctgacttt gaccttgcca 1020
gtttctgaac accatatcat cacagccaaa ctaatcggtg cctttatctg gtcattgatt 1080
agcaccgctg tatgtgctct aagtgtgttt attattctgg ctttaacagc tccagaatgg 1140
attcctcttt cttatgtgat tacatttgta gaaacacatc tccctcagat ctttcttaca 1200
ggtatatcct tectactaaa tactatttca ggaatcctct gcactaccc ggctatttcc 1260
attggacagc ttttcaatga ataccgtaca gcaactcgctg ttgcagtcta cattgggtac 1320
caaatcgtca ttggatttat tgaacttttc ttcaatctta gttctaat tctatgcaat 1380
tcaactggtg gactcaatga ccatttctat atgggagcag gtatagccat tgttgaagaa 1440
ctcatattca tagctatctt ttatctcgga acctactaca tcttgagaaa taagggttaat 1500
ttgctttaa 1509

```

&lt;210&gt; 152

&lt;211&gt; 1185

&lt;212&gt; DNA

<213> *Streptococcus pneumoniae*

&lt;400&gt; 152

```

aaaagctgtc cgcaagtgtg tccagatgtc attgacctct tggtaacacc attcgtgaca 60
cttttggtca tgtctatcct tggactcttt gtcattggac cagttttcca cgtttgtgaa 120
aactacatcc ttattgtcac aaaagcatt cttagcatgc catttggctc tgggtggttc 180
ttgattgggtg ggggttcacca attgatcgtc gtgtcagggtg tgcaccacat ctccaacttg 240
cttgaagtgc aattacttgc tgcagaccat gctaaccatc tcaacgctat catcacagct 300
gctatgacag ctcaagggtc tgctactgtt gcggttggtg ttaaaacaaa aaatccaaaa 360
ctgaaaaaac ttgctttccc ggctgctctt tctgccttcc taggtattac agagcctgct 420
atcttcgggg tgaacttgcg ctccgtaaa ccattcttcc tttcattgat tgcgtggtgca 480

```



```

atcgggtggtg gattggcttc tacccttggg cttgctggta ctggtaaatgg tatcaccatc 540
atccctggta caatgcttta tgttggtaac ggacaacttc cacaatacct tcttatggta 600
gctgtatcat ttgcccttgg ttttgccttt acttacatgt ttggttacga agatgaagta 660
gacgcaactg cagctgcaaa acgagctgaa gtggctgaag aaaaagaaga agttgcgcca 720
gcagctcttc aaaaatgaac acttgtaact cctatcgtcg gtgatgttgt cgtctctgct 780
gatgtcaatg acccagctct ctcaagtggg gctatggggc aaggatcgtg tgtgaaacca 840
agccaaggcg tggctctatgc accagctgat gctgaagttt caattgcctt tccaacaggg 900
cacgcttttg gtttgaaaaa aagaatgggt gctgaagttt tgattcatgt tgggtattgat 960
actgtatcta tgaacggtga cgggtttgaa acaaaagttg ctcaaggtaa taagggtgaa 1020
gctggcgatg ttcttggaaac atttgactca aacaaaatcg ctgcagctgg acttgatgat 1080
acaacaatgg ttatcgttac aaatacaggt gactacgctt cagtagcttc agtcgcaaca 1140
ggttcagttg ctaaggggga tgctgtgatc gaagtgaaaa tctaa 1185

```

```

<210> 153
<211> 792
<212> DNA
<213> Streptococcus pneumoniae

```

```

<400> 153
aatcgctttc aaacaagaac aaaatgttat ataaggagat ttttgcaaat gaacaatcag 60
gaaattgcaa aaaaagtcac cgatgccttg ggcggacgtg aaaatgtcaa tagtggttgc 120
cactgtgcga ctgcgtcacg tgtcatggtc aaagatgaag agaaaatcaa taaagaagtg 180
attgagaact tggaaaaagt tcaagggtct ttctttaact cagggaata ccaattatc 240
tttggatcac gtacagttaa caaaatgtac gatgaagttg ttgtacttgg attaccaaca 300
tcattctaag atgacatgaa agcagaagtt gctaaacaag ggaactggtt ccaactgtct 360
atcgttactt ttggtgatgt tttcgttcca atcatccag ttatcgtacg gacaggtctc 420
ttcatgggtg tgcgtggtct tttcaacgtc cttgaaatgc cacttcacgg tgactttgca 480
acttacacac aatcttgac agatacagcc ttcatcatct tgccaggttt ggttgtgtgg 540
tcaaccttcc gtgtatttgg tggaaacct gccgttggta tcgttcttgg tatgatgctt 600
gtctctggct cacttccaaa cgcttgggca gttgctcaag tgggtgaagt aacagcgatg 660
aaactctttg gtttcacccc tgttgttggt ttgcaaggtt ccgttcttcc agccttcac 720

```

atcgggggttg tcggagctaa atttgaaaaa gctgtccgca agttgttoca gatgtcattg 780  
acctcttggt aa 792

<210> 154  
<211> 651  
<212> DNA  
<213> Streptococcus pneumoniae

<400> 154  
acaaatcaa gaattttctg tctatttttt gaatatttat ggagaatgag actgatgaaa 60  
atatgggtata atgaataaaa ggagttttat atgcaaaaat ttattcaggc ttatattgaa 120  
aagctagatg tgacaacat tatcgagaat attctaacca aggtcatttc tcttttactg 180  
ctttaattg tattttatat tgctaaaaa atgcttcata ccattggtgca gagaattgtc 240  
aaaccttctc taaaaatgtc tcgtcatgat gttggacgcc aaaaacat ctcacgttta 300  
ctagaaaatg tgtttaatta tacgctatat ttctttttac tctactgcat tttgtcgatt 360  
ttaggtttgc cagtttctag tttgtgtggt ggagctggta ttgctggggt agcgattggt 420  
atgggagccc aaggctttct gtctgatgac atcaatgggt ttttcactct ctttgaacgt 480  
caactggatg tgggagatga ggtcgttctg acaaatggac cgattactgt atcgggtaag 540  
gttgctcagt tgggaattcg taagacacag cttcgttagc aggagcaagc ccttcacttt 600  
gtccctaacc gaaatatcac agttgttagc aattctcac gcacagacta g 651

<210> 155  
<211> 1815  
<212> DNA  
<213> Streptococcus pneumoniae

<400> 155  
agaaataaga ggaagaaaat ggaacaaaa caccgttcag aatttcaga gaaggaaactc 60  
tgggacttaa cagccctata ccaagaccgt gaggatttct tgcgtgcaat cgagaaagct 120  
cgcgaagaca tcaaccagtt tagcgtgat tacaagggca atcttcacac ttttgaggat 180  
ttcgagaagg cctttgcgga attggaacag atctacattc agatgagcca tattggcaac 240  
tatggtttta tgccctcagc gacggactat agcaatgacg aatttgcaa tattgcccac 300  
gctgggatgg aatttgaaac agatgccagc gtagccttga ccttctttga cgatgccttg 360  
gtggcagcag atgaggaagt cttggaccgt ttgggtaaat tgccacattt aacagctgcc 420

```

attcgtcagg ctaaaatcaa aaaagccac tacttagggg cagatgtgga gaaggccttg 480
acaaatctcg gtgaagtttt ctacagtcg caggacattt atactaagat gcgagctggg 540
gattttgaaa tggctgactt tgaagccat ggcaagacct acaaaaacag ctttgtgacc 600
tatgagaatt tctacaaaaa ccatgaggat gctgagggtc gtgagaaatc cttccgttcc 660
ttctcagagg gacttcgtaa gcacaaaaat acggctgcag cagcctatct ggctcaggtc 720
aagtcgaaa aactcttgcc tgatatgaag ggatacgact ctgtctttga ctatcttcta 780
gctgaacaag aagtggaccg tgtcatgttt gaccgccaga ttgacctcat catgaaggac 840
tttgaccag tcgctcagag atacctcaag catgttgcca aggtaaatgg tcttgaaaag 900
atgacctttg cagactggaa attggacttg gacagcgccc tgaatcctga agtgactatt 960
gacgatgcct atgatttggt catgaagtcg gtagaacctt tggggcaaga atattgtcag 1020
gaagttgctc gttaccaaga agagcgctgg gtggactttg ctgctaacag tggcaaggat 1080
tcoggtgggt atgcggcgga cccatatgcg gtacacctt atgtactcat gagctggaca 1140
ggcgtttga gcgatgtcta taccttgatt catgaaatcg ggcattctgg tcaattcatc 1200
ttttcagaca atcatcaaa ttacttcaat gcccatatgt cgacctacta tgttgaagca 1260
cgtcaacct tcaatgaatt gctactcagt gattacttgg agaaccagtc taatgaccca 1320
cgtaaaaac gcttcgctct ggctcatgcg ttgacagaca cctacttcca taactttatc 1380
accacacctt tggaaagccg cttccagcgt aaggtgtata cattgattga agaaggggag 1440
acctttggag caagcaagct caacagcatt atgaaggag ttttgacgga tttctgggga 1500
gatgctattg aaattgacga tgatgcaact ctgacttgga tgcgccaaag tcaactactat 1560
atgggcttgt atagttacac ttactcagca ggactagtta tctcgactgc tggttacctt 1620
catctaaaac attctgaaac tggagctgaa gactggctca atctectcaa atcaggtggt 1680
agcaagacac cacttgagtc agccatgatt atcgggtgctg atatttcaac agacaaacca 1740
ctcgtgata ccatccaatt cttgtctgac acagttgacc agattatctc ctatagtgtc 1800
gagttgggag agtag 1815

```

```

<210> 156
<211> 615
<212> DNA
<213> Streptococcus pneumoniae
<400> 156

```